2006

Association Genetics for Agronomic Traits in Rice and Cloning of ALS Herbicide Resistant Genes from Coreopsis Tinctoria Nutt

Nengyi Zhang
Louisiana State University and Agricultural and Mechanical College, nzhang1@lsu.edu

ASSOCIATION GENETICS FOR AGRONOMIC TRAITS IN RICE AND CLONING OF ALS HERBICIDE RESISTANT GENES FROM *COREOPSIS TINCTORIA* NUTT.

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

In

The Department of Agronomy and Environmental Management

By
Nengyi Zhang
B.S., Hangzhou University, 1991
M.S., Zhejiang Agricultural University, 1994
August 2006

This work

is dedicated

to my father,

Xingjian Zhang

and my mother,

Fengnu Zhang

ACKNOWLEDGEMENTS

I wish to express my profound gratitude to Dr. James Oard, my major professor, for his support, encouragement, guidance and useful suggestions throughout the period of this research. Without his help, it would be impossible for me to complete any portion of this work.

I am greatly thankful to my committee members, Drs. Gerald Myers, Don Labonte, Charles Johnson, and Seth Johnson for their contributions to my studies and research. I gratefully appreciate the contributions from our collaborators Drs. Yunbi Xu and Susan McCouch from Cornell University and Dr. Dearl Sanders from Louisiana State University Agricultural Center. I am grateful for Dr. Svetlana Oard and my brother, Dr. Nenggang Zhang for their useful discussion and suggestions.

I am greatly thankful to Louisiana State University Agricultural Center who provided the Fellowship of Biotechnology Education for Students and Teachers (BEST) for this degree and funding for the research. I want to express my sincere acknowledgement to the head and staff of the Department of Agronomy and Environmental Management for making it possible for me to study here.

I thank my colleagues, Muhanad Akash, Gabriel Aluko, Dawen Liu, Alicia Ryan, Baogong Jiang, Joohyun Lee, Weiqiang Zhang, Suresh Kadaru, Samuel Ordonez, Suman Andru, Sredhar Alwala, Niranjan Baisakh and others for their assistance.

Finally, my deepest thanks will go to my dear wife, Zhifang Cheng and our son, Jack for their love, understanding, supporting and encouragement during my study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

We have evaluated the potential of discriminant analysis (DA) to detect candidate markers associated with twelve economically important traits in a large population of unrelated U.S. and Asian inbred lines of rice. Associated marker alleles detected by DA mapped within the same genetic intervals when compared with previous traditional QTL mapping experiments that evaluated progeny derived from various controlled crosses. New markers identified by DA suggest that the procedure can also uncover relevant genetic regions not possible by standard genetic tests. With the same dataset, we also compared different modern regression approaches for selecting molecular markers associated with the twelve agronomic traits. These methods included stepwise forward regression (SFR), least angle regression (LAR) and least absolute shrinkage and selection operator (LASSO) selection. The epistatic model based on stepwise forward regression did successfully identify several interacting loci that explained a relatively high proportion of the observed variation for all the twelve agronomically important traits. Moreover, the loci identified by the epistatic model mapped within previously known QTL regions that underscores the genetic basis of the selected markers. It was concluded that stepwise forward regression with consideration for population structure, epistatic interactions, and missing data (multiple imputation) was a robust method, compared to the general linear model, to identify markers associated with complex agronomic traits.

Acetolactate synthase (ALS), also known as acetohydroxy acid synthase (AHAS), which catalyzes the first step in the biosynthesis of the branched-chain amino acids valine, leucine and isoleucine in plants, is a target of five herbicide groups, including sulfonylurea and imidazolinone. A recently discovered group of *Coreopsis tinctoria* Nutt. mutants from the field showed high levels of resistance to both sulfonylurea and imidazolinone herbicides. In this study

the mutants were compared by chemical, genetic, and molecular analyses with “normal” or wild-type *Coreopsis.* A phylogenetic analysis revealed that the ALS gene can serve as a useful molecular tool for evaluating evolutionary relationships among plant species. Due to pending patent applications by the Louisiana State University Agricultural Center and restrictions of patent applications, specific results from this research cannot be presented in this dissertation.

CHAPTER 1: INTRODUCTION

1.1 Rapid Identification of Candidate Markers Associated with Agronomic Traits among Inbred Lines of Rice Using Association Genetics

Almost all agronomically important traits such as yield and its components are quantitatively inherited showing continuous variation in segregating populations. It is assumed that quantitative traits are controlled by multiple genetic factors each having a small effect on the expression of the trait, known as the multiple factor hypothesis (East 1916). Moreover, the expression of genes controlling quantitative traits can be greatly influenced by the environment, referred to as genotype by environment (GE) interactions (Lynch and Walsh 1998). Consequently, the improvement of quantitative traits by traditional methods can be difficult, labor intensive, time consuming and costly.

Genetic markers provide breeders a potential tool to trace quantitative traits that would otherwise be impossible by conventional breeding means. Many different types of markers have been discovered, including morphological variants, protein polymorphisms, and DNA polymorphisms. Morphological markers were the first generation of markers to be used for identification and selection for quantitative traits. However, the limited number of the markers and the undesirable effects of many of the markers on the target phenotype make them difficult to effectively and extensively use these markers to study quantitatively inherited traits (Tanksley et al. 1998). Isozymes were the second generation of markers, and used successfully as markers to identify QTLs in maize (Stuber and Edwards 1986). The utility of isozymes is reduced by the numbers of such markers available (Aluko 2003; Xu 2003).

DNA-based markers have received the most attention since the first genetic map was established using restriction fragment length polymorphism (RFLP) (Botstein et al. 1980). DNA-based markers reflect genetic polymorphism at the DNA level, which result from any possible

differences existing in nucleotides. Compared with other types of genetic markers, DNA markers have almost no practical limitation in numbers, often have no direct phenotypic effect, and are unaffected by environment. DNA markers can be classified into four different types based on the method used for polymorphism detection: (1) DNA-DNA hybridization such as RFLPs; (2) PCR-based markers such as random amplified polymorphic DNA (RAPD) using arbitrary primers, or simple sequence repeat (SSR) using specific primers; (3) combining use of PCR and restriction enzyme technique such as amplified fragment length polymorphism (AFLP); and (4) single nucleotide polymorphisms (SNPs) (Xu 2003). Desirable DNA markers should meet the following requirements: detection of high frequency of polymorphism, codominance, abundance, whole genome coverage, high duplicability, suitability for high-throughput analysis and multiplexing, technical simplicity, cost effectiveness, small DNA amount requirement and user-friendly. Of all the types of DNA markers mentioned above, SSR most readily satisfies all the requirements. As estimated from a draft rice sequence, the density of SSRs in the genome is approximately one SSR per gene (Xu 2003).

Rice (*Oryza sativa* L.) is the world's most important food crop (Smidansky et al. 2003). Rice accounts for 23 percent of the world's supply of calories (Brar and Khush 2002), and more than a third of the world's population consume rice as the primary source of energy. Rice is one of the most important crops in Louisiana. In 2005, 523,739 acres of rice were grown, and the gross farm value of rice production was $225 million (Louisiana Summary 2005). Traditional breeding has reliably and significantly addressed world food sufficiency for the last century, but it has many inherent shortcomings. For example, standard practices require ~ 10 years to develop an improved variety primarily because important traits like yield are quantitative in nature and heavily influenced by the environment (Lynch and Walsh 1998). Nevertheless, total demand for

rice is expected to increase worldwide by 40% during the next 25 years (Brar and Khush 2002). It is therefore worthwhile to evaluate new approaches that will not only complement traditional breeding efforts to meet food sufficiency worldwide, but to promote the Louisiana rice industry.

To efficiently exploit DNA marker technology, the identification of specific marker alleles associated with agronomic traits is the first critical step. Most efforts in this area focus on utilizing controlled populations derived from just two parents to trace quantitative trait loci (QTL)-marker associations. The information derived from a particular set of controlled crosses, however, may prove questionable if breeders seek to use these markers on different genetic populations. Numerous studies have been conducted that identified QTLs contributing to the inheritance of quantitative traits including yield and yield component traits of the most important crop species such as tomato, maize and rice (Paterson et al. 1988; Xu 1997). Some common QTL regions in rice have been detected using different populations (Aluko 2003; Lin et al. 1996; Lu et al. 1997; Moncada et al. 2001; Xiao et al. 1998; Yu et al. 1997), but most of the results were inconsistent even with the same materials (Ishimaru et al. 2001; Li et al. 2000; Lin et al. 1996; Yagi et al. 2001; Zhuang et al. 2000). Moreover, the precision of QTL mapping is still a question because a detected interval of 5 cM by this procedure may contain more than 50 genes.

Linkage disequilibrium (LD), based on pairwise comparisons between observed and expected haplotype frequencies, is another method to identify markers associated with specific traits. LD is also known as gametic phase disequilibrium, gametic disequilibrium, and allelic association that measures the correlation between polymorphisms caused by a shared history of mutation and recombination events (Flint-Garcia et al. 2003). LD plays a central role in association analysis and has been applied in humans to identify SNPs associated with candidate genes or simply-inherited phenotypic traits (Pritchard and Przeworski 2001). Plant breeders and

geneticists are beginning to consider LD as a potential tool for crop improvement. Hansen et al. (2001) assessed the possibility of LD mapping of the bolting gene in sea beet using AFLP markers. Two markers showing significant linkage disequilibrium within the bolting gene were detected. The *dwarf8* polymorphisms in maize associated with variation in flowering time were detected by LD analysis (Thornsberry et al. 2001). Kraakman et al. (2004) employed LD mapping to identify markers associated with grain yield in barley.

In rice, Garris et al. (2003) characterized LD in the candidate region for *xa5*, a recessive gene conferring race-specific resistance to bacterial blight disease. They sampled 13 short segments from the 70-kb candidate region in 114 accessions of *Oryza sativa* L. and sequenced five additional segments from the adjacent 45-kb region in resistant accessions. Significant linkage disequilibrium was found between sites up to 100 kb apart. Population structure, admixture, selection and other factors, however, often lead to false positive associations and erroneous conclusions using this approach (Flint-Garcia et al. 2003)

An alternative method to rapidly identify candidate markers associated with agronomic traits among inbred lines of rice is proposed based on discriminant analysis (DA), a nonparametric, multivariate procedure developed by Fisher (1936). Balzarini et al. (2000) and Capdevielle et al. (2000) first proposed and evaluated DA in plants as a tool for selection of molecular markers associated with specific traits and allocation of rice breeding lines into target groups, combining information from agronomic and molecular data sets. DA involves the creation of "training samples" derived from selected individuals with contrasting phenotypic values. From DNA profiles of the selected lines, markers are identified by DA that best differentiate between training samples. An error rate, referred to as "% correct classification", is calculated to measure the ability of the markers to correctly assign individual lines to the training

samples. With high levels of correct classification, an association between marker and phenotype or agronomic trait is inferred.

DA has been used in plant research for diversity analysis of wild emmer wheat (Fahima et al. 2002), identification of drought-tolerant Kentucky bluegrass cultivars using morphological criteria (Ebdon et al. 1998) and to estimate position and effects of QTLs in simulated full and half-sib families (Gilbert and Le Roy 2003). Microarray expression profiling studies have used DA to identify genes and gene clusters associated with human diseases (DePrimo et al. 2003; Kari et al. 2003; Mendez et al. 2002; Musumarra et al. 2003) and to detect protein coding regions in genomic sequences (Zhang 1997; Zhang et al. 2002). Aluko (2003) compared the percentage correct classification with the markers selected by methods of stepwise DA and QTL mapping from a controlled cross in rice. The results showed that some common and different markers were detected by both methods, and markers selected by DA produced higher correct classification than standard QTL mapping techniques. Chapter two of this dissertation describes the identification of DNA markers associated with twelve characters of rice using the DA method (Zhang et al. 2005). Finally, the procedures described in this proposal were used recently in the Louisiana State University Agricultural Center to accurately assign unrelated sweetpotato clones using AFLP markers to groups defined by high and low dry matter content (Mcharo et al. 2004**)** and DA markers associated with virus resistance in sweetpotato (Mcharo et al. 2005).

Multiple regression (MR) analysis is also a strategy evaluated to identify markers associated with specific traits. Three QTLs with highly significant effects on multinucleate-microspore formation were identified by ANOVA and stepwise MR in diploid alfalfa (Tavoletti et al. 2000). MR of early yield on eight yield-related traits in cassava revealed harvest index, dry foliage weight and root diameter as the most important factors associated with early yield.

Furthermore, based on single-marker regression analysis, QTLs were detected for early yield and associated traits (Okogbenin and Fregene 2002). Stepwise MR was used in two maize inbred line crosses to identify nine yield QTLs, five of which were in the same regions as those identified by composite interval mapping (Kraja and Dudley 2000). Similar results were found for plant and ear height, but for grain moisture, composite interval mapping identified nearly twice as many QTLs as stepwise MR. Virk et al. (1996) used MR to identify associations between various quantitative traits and RAPD molecular markers with diverse Asian rice germplasm. In the analysis of genetic resources and adaptation in *Phytolacca dodecandra* L'Hér., 17 Ethiopian populations (249 individuals) were sampled along altitudinal gradients that varied from 1600 to 3000 m (Semagn et al. 2000). MR showed a strong association between some RAPD markers, altitude, temperature and rainfall. Kraakman et al. 2004 employed step-wise multiple linear regression to find markers associated with grain yield and stability in barley. Chapter three of this dissertation describes how the addition of an epistatic term in the MR model can improve selection of markers associated with agronomic traits in rice.

The primary research objective in this area was to assess the ability of DA, coupled with other procedures described here, to identify candidate markers associated with 12 agronomic traits among US and Asian rice inbred lines. Different training samples were created for each trait and the corresponding % correct classification was determined. The potential genetic basis of the DA-selected markers was evaluated by comparing their genetic map locations with QTL markers previously identified by traditional mapping approaches**.** The DA was also compared with a modified MR method to select markers associated with agronomic traits. The information obtained from this research may be useful in future rice improvement work for selection of

parents in crosses and marker-assisted selection schemes not possible by traditional breeding methods.

1.2 Genetic Studies of Mutants of *Coreopsis tinctoria* Nutt.

Weeds are a constant limitation to optimal commercial crop production and can cause substantial yield losses in all growing seasons in the world. The use of herbicides is considered an effective, easy and comparatively inexpensive approach to control noxious weeds. Acetolactate synthase (ALS) (acetohydroxyacid synthase, AHAS, E.C. 4.1.3.18), which catalyses the first common step in the biosynthesis of the branched-chain amino acids in plants, is a target of five herbicide groups, *viz.* sulfonylurea, imidazolinone, triazolopyrimidine, pyrimidinyl oxybenzoate, and sulfonylanminocarbonyl-triazolinones (Mallory-Smith and Retzinger, 2003). These herbicides block the biosynthesis of valine, leucine and isoleucine (Ray 1984; Santel et al. 1999; Schloss 1990; Shaner et al. 1984; Sibony and Rubin 2003; Stidham and Shaner 1990; Subrananian and Gerwick 1989; Subrananian et al. 1990). ALS-inhibiting herbicides have been used as selective agents in laboratory studies to isolate a range of resistant biotypes from otherwise susceptible populations (Falco and Dumas 1985; Haughn and Somerville 1986). More than 70 field plants throughout the world also have been reported showing resistance after repeated applications of ALS-inhibiting herbicides for more than three years (Heap 2003; Saari et al. 1994).

Resistance in many cases has been attributed to single point mutations which can occur at multiple sites within the ALS gene (Shaner 1999) that provide opportunities to study the molecular basis of resistance and to transfer cloned resistance genes to different economic crops for weed management. Most ALS mutant lines either from laboratory or field sources generally possess a nucleotide base-pair substitution at only one or two sites (Boutsalis et al. 1999) and do

not show a broad spectrum of resistance to ALS-inhibiting herbicides. Consequently, it would be beneficial to discover new gene sources of ALS that exhibit a broad spectrum of resistance for basic biochemical and molecular studies and to transfer high levels of resistance to commercial crops. Dr. Dearl Sanders of the Louisiana State University Agricultural Center recently discovered *Coreopsis tinctoria* Nutt. plants along a highway in Louisiana that are resistant to ALS and AHAS inhibiting herbicides (http://www.lsuagcenter.com/inst/research/stations/idlewild/pdfs/WeedScience/Wildflower1.pdf). Unlike most other ALS mutants produced from chemical mutagenesis, the *Coreopsis tinctoria* Nutt. plants are highly resistant to both sulfonylurea (sulfometuron methyl "Oust®"; chlorsulfuron "Glean®") and imidazolinone (imazapyr "Arsenal®"; imazapic "Plateau®") herbicides.

Two parallel reactions are catalyzed by the ALS enzyme: synthesis of (S)-2-acetolactate from two molecules of pyruvate and synthesis of (S)-2-aceto-2-hydroxybutyrate from a molecule each of pyruvate and 2-ketobutyrate (Guttieri et al. 1996; McCourt et al. 2006; Singh et al. 1988). (S)-2-acetolactate is a precursor of valine and leucine while (S)-2-aceto-2-hydroxybutyrate is a precursor of isoleucine. In eukaryotes, ALS is encoded in the nucleus and is located in plastids of plants (Bowen et al. 1997; Duggleby and Pang 2000) or in mitochondria of fungi (Duggleby and Pang 2000). An N-terminal transit peptide is presumed to direct the protein to the appropriate organelle, and it is usually assumed that this transit peptide is cleaved during or after translocation. The site of cleavage has not yet been established for any ALS protein (Duggleby and Pang 2000). Most diploid plant species have a single ALS locus (*Arabidopsis thaliana* (L.) Heynh. and *Xanthium strumarium* L.), with corn (*Zea mays* L.) and sunflower (*Helianthus annuus* L.*)* being notable exceptions with two loci (Guttieri et al. 1996) and three loci (Kolkman et al. 2004), respectively. Tetraploid tobacco (*Nicotiana tabacum* L.) has two loci

(Chaleff and Bascomb 1987); *Brassica* species possess five loci (Rutledge et al. 1991); and *Gossypium Hirsutum* L. contain six loci (Grula et al., 1995). *Coreopsis tinctoria* Nutt. is reported to be a diploid species (Strother 1983), but the exact number of ALS loci is unknown. The mature ALS protein is approximately 670 amino acids long (Tan et al. 2005) and is highly conserved across species (Guttieri et al. 1996).

Certain related classes of herbicides are known to inhibit ALS, such as the sulfonylurea, imidazolinone, triazolopyrimidine, pyrimidinyl oxybenzoate, and sulfonylanminocarbonyl-triazolinones, by binding to a relic quinine-binding site (Hattori et al. 1995, Santel et al. 1999). A major advantage of these compounds is their high efficacy, broad-spectrum weed control, low use rates, and environmental safety (Mazur and Falco 1989). Chlorsulfuron (Glean®) was the first ALS inhibitor marketed in North America in 1982 (Guttieri et al. 1996). Sulfometuron methyl (Oust®) is another sulfonylurea herbicide widely used (Guttieri et al. 1996). Imidazolinone herbicides, such as imazethapyr (Pursuit®) and imazaquin (Scepter®) are broad-spectrum weed control herbicides (Guttieri et al. 1996). Imazapic (Plateau®) is a selective herbicide for both the pre and post-emergent control of grasses, broad-leaf weeds, and weed control in natural areas, particularly in conjunction with the establishment of native warm-season prairiegrasses and certain legumes. Imazapic is relatively non-toxic to terrestrial and aquatic mammals, birds, and amphibians. Imazapic has an average half-life of 120 days in soil, and is rapidly degraded by sunlight in aqueous solution (Tu et al. 2001). Imazapyr (Arsenal®) is a non-selective herbicide used for the control of a broad range of weeds including annual and perennial grasses and broadleaved herbs, and woody species. Imazapyr is not highly toxic to birds and mammals, but some formulations can cause severe, irreversible eye damage. Studies indicate imazapyr is excreted by mammalian systems rapidly with no bioaccumulation (Tu et al. 2001).

Worldwide, there are more than 30 commercial ALS-inhibiting herbicides, indicative of their importance for weed management in a wide range of crops (Shaner 1999).

The lack of inhibition of ALS in resistant plant biotypes is predominantly due to an altered form of ALS that is insensitive to certain herbicides (Christopher et al. 1992; Devine et al. 1991; Manley et al. 1999; Saari et al. 1990, 1992, 1994; Thill et al. 1993). A second mechanism of resistance is enhanced herbicide metabolism resulting in rapid detoxification of the herbicide (Christopher et al. 1992; Saari et al. 1994; Veldhuis et al. 2000). Examples for the second mechanism include primisulfuron-tolerant corn (*Zea mays* L.) (Guttieri et al. 1996), chlorsulfuron-tolerant soybean (*Glycine max* (L.) Merr.) (Guttieri et al. 1996), rigid ryegrass (*Lolium rigidum* Gaud.) (Christopher et al. 1991; Cotterman and Saari 1992; Holtum et al. 1991), and blackgrass (*Alopecurus myosuroides* Huds.) (Kemp et al. 1990; Moss and Cussans 1991).

Single point mutations within multiple sites of the ALS gene can result in a variable pattern of cross-resistance between the classes of ALS-inhibiting herbicides (Shaner 1999). Initially these point mutations were characterized using mutants generated in the laboratory, e.g., ALS-inhibiting herbicide resistant tobacco (*Nicotiana tabacum* L.) cell cultures (Chaleff and Ray 1984; Creason and Chaleff 1988; Hartnett et al. 1990; Lee et al. 1988), *Arabidopsis thaliana* (L.) Heynh. seeds (Haughn and Somerville 1986, 1990; Haughn et al. 1988; Mourad and King et al. 1992; Mourad et al. 1993; Sathasivan et al. 1991), corn (*Zea mays* L.) cultures (Bernasconi et al. 1995), and *Brassica napus* L. cell cultures (Hattori et al. 1995). The same mutations were later detected in field-resistant plants, e.g., lettuce (*Lactuca serriola* L.) (Guttieri et al. 1992), kochia (*Kochia scoparia* (L.) Schrad.) (Guttieri et al. 1995; Saari et al. 1990), cocklebur (*Xanthium strumarium* L.) (Bernasconi et al. 1995), *Raphanus raphanistrum* L. (Boutsalis 2001; Hashem et al. 2001; Hashem and Bowran 2002; Tan and Medd 2002; Walsh et al. 2001), *Lindernia* (Itoh

and Wang 1997; Itoh et al. 1999; Uchino and Watanabe 2002; Uchino et al. 1999, 2000), and *Amaranthus blitoides* S. Watson (Sibony and Rubin 1996, 2003). Biotypes in at least 20 monocotyledonous and 44 dicotyledonous plant species were recorded as having evolved resistance to several of the ALS-inhibiting herbicides (Heap 2003). This may be due to repeated applications of ALS inhibitor herbicides for more than three years (Rubin 1996).

The majority of mutations known to confer resistance to imidazolinone herbicides have been detected in domains A and B of the large subunit of the ALS gene (Tan et al. 2005; Wright et al. 1998). These mutations occur at positions (codons) Ala122, Pro197, Ala205, Trp574, and Ser653 (Amino acids numbered according to *Arabidopsis thaliana* (L.) Heynh. described in Sathasivan et al. 1990). Mutations at Ala122, Ser653, and Ala205 generally confer resistance to imidazolinones, but not sulfonlyureas (Tan et al. 2005). The most common mutations in biotypes selected by sulfonylureas occur in the highly conserved domain A with 13 amino acids, where any alteration of Pro197 confers resistance primarily to sulfonylureas and triazolopyrimiines (Guttieri et al. 1992). A Trp574 to Leu mutation in domain B has been associated with broad cross-resistance to representatives of all five families of ALS-inhibiting chemicals (Bernasconi et al. 1995; Tranel et al. 2006; Woodworth et al. 1996b). Some mutations occur in domain C where an Ala122 to Thr mutation appears to confer resistance only to imidazolinones (Bernasconi et al. 1995), while an Ala205 to Val substitution in domain D confers broad cross-resistance (Woodworth et al. 1996a), as in the case of the Trp574 codon in Domain B. In nearly all instances of enzyme-based resistance to ALS herbicides, resistance has been inherited as a single gene with varying degrees of dominance (Tranel and Wright 2002). Currie et al. (1995) demonstrated that ALS extracts from Pioneer IR corn hybrids were 6-fold more resistant to imazethapyr when compared to more than 62-fold resistance in homozygous plants. In the

heterozygous XI-12 corn, imazethapyr resistance was 5-fold, compared to 250-fold in the homozygous plants, indicating that resistance in corn XI-12 is a semidominant trait (Wright and Penner 1998). Similar results were also obtained in *Sisymbrium orientale* L. (Boutsalis et al. 1999).

The objective of this research is to conduct genetic and molecular analysis of the *Coreopsis tinctoria* Nutt. mutants. Specific goals and research results will not be presented here due to pending patent applications by the Louisiana State University Agricultural Center.

1.3 References

Aluko G.K. 2003. Genetic mapping of agronomic traits from the interspecific cross *Oryza sativa* (L.) and *Oryza glaberrima* (Steud.). Ph.D. dissertation, Louisiana State university.

Balzarini M., Aluko G.K., Capdevielle F.M., and Oard J.H. 2000. Towards allocation of rice lines into predefined groups: discriminant analysis using molecular marker information. In: Procedings of the 28th Rice Technical Working Group, Biloxi, MS. pp51

Bernasconi P., Woodworth A.R., Rosen B.A., Subramanian M.V., and Siehl D.L. 1995. A naturally occurring point mutation confers broad range tolerance to herbicides that target acetolactate synthase. J. Biol. Chem. 270:17381-17385

Botstein D., White R.L., Skolnick M., and Davis R.W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32:314-331

Boutsalis P. 2001. Syngenta Quick-Test: a rapid whole-plant test for herbicide resistance. Weed Tech. 15:257-263

Boutsalis P., Karotam J., and Powles S.B. 1999. Molecular basis of resistance to acetolactate synthase-inhibiting herbicides in *Sisymbrium orientale* and Brassica *tournefortii*. Pesticide Sci. 55:507-516

Bowen T.L., Union J., Tumbula D.L., and Whitman W.B. 1997. Cloning and phylogenetic analysis of the genes encoding acetohydroxyacid synthase from the archaeon *Methanococcus aeolicus*. Gene 188:77-87

Brar D.S. and Khush G.S. 2002. Transferring genes from wild species into rice. In: Kang M.S. (ed) Quantitative genetics, genomics and plant breeding. CABI Publishing, New York, USA, pp197-217

Capdevielle F.M., Aludo G.K., Balzarini M., and Oard J.H. 2000. Application of molecular markers and discriminant analysis to identify rice lines with contrasting phenotypes for agronomic traits. In Fourth International Rice Genetics Symposium, International Rice Research Institute, Philippines, pp216

Chaleff R.S. and Bascomb N.F. 1987. Genetic and biochemical evidence for multiple forms of actolactate synthase in *Nicotiana tabacum*. Mol. Gen. Genet. 210:33-38

Chaleff R.S. and Ray T.B. 1984. Herbicide-resistant mutants from tobacco cell cultures. Science 223:1148-1151

Christopher J.T., Powles S.B., and Holtum J.A.M. 1992. Resistance to acetolactate synthase-inhibiting herbicides in annual ryegrass (*Lolium rigidum*) involves at least two mechanisms. Plant Physiol. 100:1909-1913

Christopher J.T., Powles S.B., Liljegren D.R., and Holtum J.A.M. 1991, Cross-resistance to herbicides in annual ryegrass (*Lolium rigidum*). Plant Physiol. 95:1036–1043

Cotterman J.C. and Saari L.L. 1992. Rapid metabolic inactivation is the basis for cross-resistance to chlorsulfuron in diclofop-methyl–resistant rigid ryegrass (*Lolium rigidum*) biotype SR4/84. Pestic. Biochem. Physiol. 43:182–192

Creason G.L. and Chaleff R.S. 1988. A second mutation enhances resistance of a tobacco mutant to sulfonylurea herbicides. Theor. Appl. Genet. 76:177-182

Currie R.S., Kwon C.S., and Penner D. 1995. Magnitude of imazethapyr resistance of corn (Zea mays) hybrids with altered acetolactate synthase. Weed Sci. 43:578-582

DePrimo S.E., Wong L.M., Khatry D.B., Nicholas S.L., Manning W.C., Smolich B.D., O'Farrell A.-M., and Cherrington J.M. 2003. Expression profiling of blood samples from an SU5416 Phase III metastatic colorectal cancer clinical trial: a novel strategy for biomarker identification. BMC Cancer 3:1-12

Devine M.D., Marles M.A.S., and Hall L.M. 1991. Inhibition of acetolactate synthase in susceptible weed biotypes. Weed Sci. 40:670-676

Duggleby R.G. and Pang S.S. 2000. Acetohydroxyacid synthase. J. Biochem and Mol. Biol. 33:1-36

East E.M. 1916. Studies on size inheritance in *Nicotiana*. Genetics 1:164–176

Ebdon J.S., Petrovic A.M., and Schwager S.J. 1998. Evaluation of discriminant analysis in identification of low-and-high-water use Kentucky bluegrass cultivars. Crop Sci. 38:152-157

Fahima T., Roder M.S., Wendehake K., Kirzhner V.M., and Nevo E. 2002. Microsatellite polymorphism in natural populations of wild emmer wheat, Triticum dicoccoides, in Israel. Theor. Appl. Genet 104:17-29

Falco S.C. and Dumas K.S. 1985. Genetic analysis of mutants of *Saccharomyces cerevisiae* resistant to the herbicide sulfometuron methyl. Genetics 109:21-35

Fisher R.A. 1936. The use of multiple measurements in taxonomic problems. Ann Eugenics 7:179-188

Flint-Garcia S.A., Thornsberry J.M., and Buckler E.S. 2003. Structure of linkage disequilibrium in plants. Ann. Rev. Plant Biol. 54:357-374

Garris A.J., McCouch S.R., and Kresovich S. 2003. Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the xa5 locus of rice (*Oryza sativa* L.). Genetics 165:759-769

Gilbert H. and Le Roy P. 2003. Comparison of three multitrait methods for QTL detection. Genet. Sel. Evol. 35:281-304

Grula J.W., Hudspeth R.L., Hobbs S.L., and Anderson D.M. 1995. Organization, inheritance and expressionof acetohydroxyacid synthase genes in the cotton allotetraploid *Gossypium hirsutum*. Plant Molec. Biol. 28:837-846

Guttieri M.J., Eberlein C.V., and Thill D.C. 1995. Diverse mutations in the acetolactate synthase gene confer chlorsulfuron resistance in kochia (*Kochia scoparia*) biotypes. Weed Sci. 43:175-182

Guttieri M.J., Eberlein C.V., Mallory-Smith C.A., and Thill D.C. 1996. Molecular genetics of target site resistance to acetolactate synthase inhibiting herbicides. In Molecular Genetics and ecology of Pesticide Resistance, T.M. Brown, ed (Washington, DC: American Chemical Society). pp10-16

Guttieri M.J., Eberlein C.V., Mallory-Smith C.A., Thill D.C., and Hoffman D.L. 1992. DNA sequence variation in domain A of the acetolactate synthase genes of herbicide resistant and susceptible weed biotypes. Weed Sci. 40:670-676

Hansen M., Kraft T., Ganestam S., Sall T., and Nilsson N.O. 2001. Linkage disequilibrium mapping of the bolting gene in sea beet using AFLP markers. Genet. Res., Camb. 77:61-66

Hartnett M.E., Chui C.F., Mauvais C.J., McDevitt R.E., Knowlton S., Smith J.K., Falco S.C., and Mazur B.J. 1990. Herbicide-resistant plants carrying mutated acetolactate synthase genes. In managing resistance to agrochemicals, from fundamental research to practical strategies. M.B. Green, H.M. LeBaron, and W.K. Moberg, eds. ACS Symp. Ser. Am. Chem. Soc. 421, pp459-473

Hashem A. and Bowran D. 2002. Seed soaking method to test resistance in *Raphanus raphanistrum* to ALS-inhibiting herbicides. Plant Physiol. 92:1081-1085

Hashem A., Bowran D., Piper T., and Dhammu H. 2001. Resistance. of target site resistance to acetolactate synthase inhibiting herbicides in the Western Australia wheat belt. Weed Tech. 15:68-74

Hattori J., Brown D.C., and Miki B. 1995. An acetohydroxy acid synthase mutant reveals a single site involved in multiple herbicide resistance. Mol. Gen. Genet. 246:419-425

Haughn G.W. and Somerville C. 1986. Sulfonylurea-resistant mutants of *Arabidopsis thaliana*. Mol. Gen. Genet. 204:430-434

Haughn G.W. and Somerville C.R. 1990. A mutation causing imidazolinone resistance maps to *csr1* locus in *Arabidopsis thaliana* var Columbia. Plant Physiol. 92:1081-1085

Haughn G.W., Smith J., Mazur B., and Somerville C. 1988. Transformation with a mutant Arabidopsis acetolactate synthase gene renders tobacco resistant to sulfonylurea herbicides. Mol. Gen. Genet. 211:266-271

Heap I. 2003. International survey of herbicide resistant weeds. http://www.weedscience.com

Holtum J.A.M., Matthews J.M., Häusler R.E., Liljegren D.R., and Powles S.B. 1991. Cross resistance to herbicides in annual ryegrass (*Lolium rigidum*). III. On the mechanism of resistance to diclofop-methyl. Plant Physiol. 97:1026–1034.

Ishimaru K., Yano M., Aoki N., Ono K., Hirose T. S., Lin Y., Monna L., Sasaki T., and Ohsugi R. 2001. Toward the mapping of physiological and agronomic characters on a rice function map: QTL analysis and comparison between QTLs and expressed sequence tags. Theor. Appl. Genet. 102:793–800

Itoh K. and Wang G.-X. 1997. An outbreak of sulfonylurea herbicide resistance in Scrophulariaceae paddy weeds in Japan. In Proceedings of the 16th Asian-Pacific Weed Science Socity Conference (Kuala Lumpur, Malaysia, 8-12 September 1997) (Kuala Lumpur: Malaysian Plant Protection Society), pp219-221

Itoh K., Wang G.-X., and Ohba S. 1999. Sulfoylurea resistance in *Lindernia micrantha*, an annual paddy weed in Japan. Weed Res. 39:413-423

Kari L., Loboda A., Nebozhyn M., Rook A.H., Vonderheid E.C., Nichols C., Virok D., Chang C., Horng W.-H., Johnston J., Wysocka M., Showe M.K., and Shwe L.C. 2003. Classification and prediction of survival in patients with the leukemic phase of cutaneous T cell lymphoma. J. Exp. Med. 197:1477-1488

Kemp M.S., Moss S.R., and Thomas T.H. 1990. Herbicide resistance in Alopecurus myosuroides. In: Green M.B., LeBaron H.M., and Moberg W.K. (eds) Managing resistance to

agrochemicals. From fundamental research to practical strategies. Washington D.C.: American Chemical Society. pp376-393

Kolkman J.M., Slabaugh M.B., Bruniard J.M., Berry S., Bushman B.S., Olungu C., Maes N., Abratti G., Zambelli A., Miller J.F., Leon A., and Knapp S.J. 2004. Acetohydroxyacid synthase mutations conferring resistance to imidazolinone or sulfonylurea herbicides in sunflower. Theor. Appl. Genet. 109:1147-1159

Kraakman A.T.W., Niks R.E., Van den Berg P.M.M.M., Stam P., and Van Eeuwijk F.A. 2004. Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics 168:435-446

Kraja A.T. and Dudley J.W. 2000. QTL analysis of two maize inbred line crosses. Maydica 45:1-12

Lee K.Y., Townsend J., Tepperman J., Black M., Chui C.F., Mazur B., Dunsmuir P., and Bedbrook J. 1988. The molecular basis of sulfonylurea herbicide resistance in tobacco. EMBOJ 7:1241-1248

Li J.X., Yu S.B., Xu C.G., Tan Y.F., Gao Y.J., Li X.H., and Zhang Q. 2000. Analyzing quantitative trait loci for yield using a vegetatively replicated F inferior 2 population from a cross between the parents of an elite rice hybrid. Theor. Appl. Genet. 101:248-254

Lin H.X., Qian H.R., Zhuang J.Y., Lu J., Min S.K., Xiong Z.M., Huang N., and Zheng K.L. 1996. RFLP mapping of QTLs for yield and related characters in rice (*Oryza sativa L.*) Theor. Appl. Genet. 92:920-927

Louisiana Summary 2005. http://www.lsuagcenter.com/agsummary/narrative.aspx#Rice. Visited on June 1, 2006

Lu C., Shen L., Tan Z., Xu Y., He P., Chen Y., and Zhu L. 1997. Comparative mapping of QTLs for agronomic traits of rice across environments by using a doubled-haploid population. Theor. Appl. Genet. 94:145–150

Lynch M. and Walsh B. 1998. Genetics and analysis of quantitative traits. Sinauer Associates, Sunderland, MA.

Mallory-Smith C.A. and Retzinger E.J. 2003. Revised classification of herbicides by site of action of weed resistance management strategies. Weed Tech. 17:605-619

Manley B.S., Singh B.K., Shaner D.L., and Wilson H.P. 1999. Imidazolinone resistance in smooth pigweed (*Amaranthus hybridus*) is due to an altered acetolactate synthase. Weed Tech. 13:697-705

Mazur B.J. and Falco S.C. 1989. The development of herbicide resistant crops. Ann. Rev. Plant Physiol. Plant Mol. Biol. 40:441-470

McCourt J.A., Pang S.S., King-Scott J., Guddat L.W., and Duggleby R.G. 2006. Herbicide-binding sites revealed in the structure of plant acetohydroxyacid synthase. PNAS 103:569-573

Mcharo M., LaBonte D.R., Clark C., Hoy M., and Oard J. H. 2005. Molecular marker variability for southern root-knot nematode resistance in sweetpotato. Euphytica 144:125-132

Mcharo M., Labonte D.R., Oard J.H., Kays S.J., and McLaurin W.J.. 2004. Linking quantitative traits with AFLP markers in sweetpotato using Discriminant Analysis. Acta Horticultura 637:285-293

Mendez M.A., Hodar C., Vulpe C., Gonzalez M., and Cambiazo V. 2002. Discriminant analysis to evaluate clustering of gene expression data. FEBS Letters 522:24-28

Moncada P., Martínez C.P., Borrero J., Chatel M., Gauch Jr H., Guimaraes E., Tohme J., and McCouch S.R. 2001. Quantitative trait loci for yield and yield components in an *Oryza sativa* × *Oryza rufipogon* BC2F2 population evaluated in an upland environment. Theor. Appl. Genet. 102:41–52

Moss S.R. and Cussans G.W. 1991. The development of herbicide resistant populations of *Alopecurus myosuroides* (black-grass) in England. In Caseley J.C., Cussans G.W., and Atkin R.K. (eds). Herbicide resistance in weeds and crops. Oxford, U.K.: Butterworth-Heneman pp445-456

Mourad G. and King J. 1992. Effect of four classes of herbicides on growth and acetolactate-synthase activity in several variants of *Arabidopsis thaliana*. Planta 188:491-497

Mourad G., Pandey B., and King J. 1993. Isolation and genetic analysis of a triazolopyrimidine-resistant mutant of *Arabidopsis*. *J. Hered*. 84:91-96

Musumarra G., Barresi V., Condorelli D.F., and Scirè S. 2003. A bioinformatic approach to the identification of candidate genes for the development of new cancer diagnostics. Biological Chemistry 384:321-327

Okogbenin E. and Fregene M. 2002. Genetic analysis and QTL mapping of early root bulking in an F_1 population of non-inbred parents in cassava (*Manihot esculenta* Crantz). Theor. Appl. Genet. 106:58–66

Paterson A.H., Lander E.S., Had J.D., Patrson S., Lincoln S.E., and Tanksley S.D. 1988. Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721–726

Pritchard J.K. and Przeworski M. 2001. Linkage disequilibrium in humans: models and data. Am. J. Hum. Genet. 69:1-14

Ray T.B. 1984. Site of action of chlorsulfuron: inhibition of valine and isoleucine biosynthesis in plants. Plant Physiol. 75:827-831

Rubin B. 1996. Herbicide-resistant weeds — the inevitable phenomenon: mechanisms, distribution and significance. Z Pflanzenkr Pflanzenschutz XV, 17-32

Rutledge R., Ouellet T., Hattori J., and Miki B.L. 1991. Molecular characterization and genetic origin of the *Brassica napus* acetohydroxyacid synthase multigene family. Mol. Gen. Genet. 229:31-40

Saari L.L., Cotterman J.C., and Primiani M.M. 1990. Mechanism of sulfonylurea herbicide resistance in the broadleaf weed, *Kochia scoparia*.. Plant Physiol. 93, 55-61

Saari L.L., Cotterman J.C., and Thill D.C. 1994. Resistance to acetolactate synthase inhibiting herbicides. In Herbicide Resistance in Plants: Biology and Biochemistry, S.B. Powles and J.A.M. Holtum, eds (Boca Raton, FL: CRC Press), pp83-139

Saari L.L., Cotterman J.C., Smith W.F., and Primiani M.M. 1992. Sulfonylurea herbicide resistance in common chickweed, perennial ryegrass, and Russian thistle, Pestic. Biochem. Physiol. 42:110-118

Santel H.J., Bowden B.A., Sorenson V..M., Mueller K.H., and Reynolds J. 1999. Flucarbazone-sodium—a new herbicide resistance in wheat. Proc. West. Soc. Weed Sci. 52:124–125

Sathasivan K., Haughn G.W., and Murai N. 1990. Nucleotide sequence of a mutant acetohydroxyacid synthase gene from an imidazolinone-resistant *Arabidopsis thaliana* var. Columbia. Nucleic Acids Res. 18:2188

Sathasivan K., Haughn G.W., and Murai N. 1991. Molecular basis of imidazolinone herbicide resistance in *Arabidopsis thaliana* var Columbia. Plant Physiol. 97:1044-1050

Schloss J.V. 1990. Acetolactate synthase, mechanism of action and its herbicide binding site. Pestic. Sci. 29:283-292

Semagn K., Bjornstad A., Stedje B., and Bekele E. 2000. Comparison of multivariate methods for the analysis of genetic resource and adaptation in *Phytolacca dodecandra* using RAPD. Theor. Appl. Genet. 101:1145-1154

Shaner D.L. 1999. Resistance to acetolactate synthase (ALS) inhibitors in the United States: history, occurrence, detection and management. J. Weed Sci. Technol. 44:405-411.

Shaner D.L., Anderson P.C., and Stidham M.A. 1984. Imidazolinones: potent inhibitors of acetohydroxyacid synthase.. Plant Phsiol. 76:545-546

Sibony M. and Rubin B. 1996. Multiple resistance to ALS inhibitors and triazine herbicides in *Amaranthus* species. In Proceedings of the international symposium on weed and crop resistance to herbicides (Cordoba, Spain), pp112-114

Sibony M..and Rubin B. 2003. Molecular basis for multiple resistance to acetolactate synthase-inhibiting herbicides and atrazine in *Amaranthus blitoides* (prostrate pigweed). Planta 216:1022-1027

Singh B.K., Stidham M.A., and Shaner D.L. 1988. Assay. of acetohydroxyacid synthase. Analytical Biochemistry 171:173-179

Smidansky E.D., Martin J.M., Hannah L.C., Fischer A.M., and Giroux M.J. 2003. Seed yield and plant biomass increases in rice are conferred by deregulation of endosperm ADP-glucose pyrophosphorylase. Planta 216:656–664

Stidham M.A. and Shaner D.L. 1990. Imidazolinone inhibition of acetohydroxyacid synthase in vitro and in vivo. Pestic. Sci. 29:335-340

Strother, J.L. 1983. More chromosome studies in Compositae. Amer. J. Bot. 70:1217-1224.

Stuber C.W. and Edwards M.D. 1986. Genotypic selection for improvement of quantitative traits in maize using molecular marker loci. In: Proceedings of the 41st Annual Maize and Sorghum Research Conference American Seed Trade Association, Chicago, Illinois, pp40-83

Subramanian M.V. and Gerwick B.C. 1989. Inhibition of acetolac- tate synthase by triazolopyrimidines. *In* J.R. Whitaker et. al. (ed.) Biocatalysis in agricultural biotechnology. ACS Symposium Series. American Chemical Society, Washington, DC. pp115-128

Subramanian M.V., Hung H.Y., Dias J.M., Miner V.W., Butler J.H., and Jachetta J.J. 1990. Properties of mutant acetolactate synthases resistant to triazolopyrimidine sulfonanilide. Plant Physiol. 94:239– 244

Tan M.K. and Medd R.W. 2002. Characterization of the acetolactate synthase (ALS) gene of *Raphanus raphanistrum* L. and the molecular assay of mutations associated with herbicide resistance. Plant Science 163:195-205

Tan S., Evans R.R., Dahmer M.L., Singh B.K., and Shaner D.L. 2005. Imidazolinone-tolerant crops: history, current status and future. Pest Manag. Sci. 61:246-257

Tanksley S.D., Bernachi D., and Beck-Bunn T. 1998. Yield and quality evaluations on a pair of processing tomato lines nearly isogenic for the Tm2(a) gene for resistance to the tobacco mosaic virus. Euphytica 99:77-83

Tavoletti S., Pesaresi P., Barcaccia G., and Albertini E. 2000. Mapping the *jp* (jumbo pollen) gene and QTLs involved in multinucleate microspore formation in diploid alfalfa. Theor. Appl. Genet. 101:372-378

Thill D.C., Mallory-Smith C.A., Saari L.L., Cotterman J.C., Primiani M.M., and Saladini J.L. 1993. Sulfonylurea herbicide resistant weeds: discovery, distribution, biology,

mechanism, and management. In Herbicide Resistance in Weeds and Crops, J.C. Caseley, G.W. Cussans, and R.K. Atkin, eds (London: Butterworth and Heinemann), pp115-128

Thornsberry J.M., Goodman M.M., Doebley J., Kresovich S., Nielsen D., and Buckler E.S. 2001. *Dwarf8* polymorphisms associate with variation in flowering time. Nature Genetics 28:286-289

Tranel P.J. and Wright T.R. 2002. Resistance of weeds to AHAS-inhibiting herbicides: what have we learned? Weed Sci. 50: 700-712

Tranel P.J., Wright T.R,, and Heap, I.M. 2006. ALS mutations from herbicide-resistant weeds. Online. Internet. Sunday, March 12, 2006. Available http://www.weedscience.com.

Tu M., Hurd C., and Randall J.M. 2001. Weed Control Methods Handbook: Tools and Techniques for Use in Natural Areas. In: http://tncweeds.ucdavis.edu/handbook.html

Uchino A. and Medd R.W. 2002. Mutations in the acetolactate synthase genes of sulfonylurea-resistant biotypes of *Lindernia* spp. Weed Biology and Management 2:104-109

Uchino A., Itoh K., Wang G.-X., and Tachibana M. 2000. Sulfonylurea resistant biotypes of *Lindernia* species in the Tohoku region and their response to several herbicides. J. Weed Sci. Technol. 45:13-20

Uchino A., Watanabe H., Wang G.-X., and Itoh K. 1999. Light requirement in rapid diagnosis of sulfonylurea-resistant weeds of *Lindernia* spp. (Scrophulariaceae). Weed Technol. 13:680-684

Veldhuis L.J., Hall L.M., O'Donovan J.T., Dyer W., and Hall J.C. 2000. Metabolism-based resistance of a wild mustard (*Sinapis arvensis* L.) biotype to ethametsulfuron-methyl. J. Agric. Food Chem. 48:2986-2990

Virk P.S., Ford-Lloyd B.V., Jackson M.T., Pooni H.S., Clemeno T.P., and Newbury H.J. 1996. Predicting quantitative variation within rice germplasm using molecular markers. Heredity 76:296-304

Walsh M.J., Duane R.D., and Powles S.B. 2001. High frequency of chlorsulfuron resistant wild radish (*Raphanus raphanistrum* L.) populations across the Western Australian wheatbelt. Weed Tech. 15:199-203

Woodworth A.R., Bernasconi P., Subramanian M.V., and Rosen B.A. 1996a. A second naturally occurring point mutation confers broad-based tolerance to acetohydroxyacid synthase inhibitors. Plant Physiol. 111S:415 (Abstract)

Woodworth A.R., Rosen B.A., and Bernasconi P. 1996b. Broad range resistance to herbicides targeting acetohydroxyacid synthase (ALS) in a field isolate of *Amaranthus* sp. is conferred by

a Try to Leu mutation in the ALS gene (Accession Number U55852) (PGR96-051). Plant Physiol. 111:1353

Wright T.R. and Penner D. 1998. Corn (Zea mays) acetolactate synthase sensitivity to four classes of ALS-inhibiting herbicides. Weed Sci. 46:8-12

Wright T.R., Bascomb N.F., Sturner S.F., and Penner D. 1998. Biochemical mechanism and molecular basis for ALS-inhibiting herbicide resistance in sugarbeet (*Beta vulgaris*) somatic cell selections. Weed Sci. 46:13-23

Xiao J.H., Li J., Grandillo S., Sang N.A., Yuan L., Tanksley S.D., and McCouch S.R. 1998. Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. Genetics 150: 899-909

Xu Y. 1997. Quantitative trait loci: separating, pyramiding, and cloning. Plant Breed Reviews 15:85–139

Xu Y. 2003. Developing marker-assisted selection strategies for breeding hybrid rice.1997. Quantitative trait loci: separating, pyramiding, and cloning. Plant Breeding Reviews 23:73-174

Yagi T., Nagata K., FukutaXu, Y., Tamura K., Ashikawa I., and Terao T. 2001. QTL mapping of spikelet number in rice (*Oryza sativa* L.). 2003. Developing marker-assisted selection strategies for breeding hybrid rice. Plant Breeding Science 51:53-56

Yu S.B., Li J.X., Xu C.G., Tan Y.F., Gao Y.J., Li X.H., Zhang Q., and Shagai Maroof M.A. 1997. Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. Proc. Natl. Acad. Sci. 94:9226–9231

Zhang C.T., Wang J., and Zhang R. 2002. Using a Euclid distance discriminant method to find protein coding genes in the yeast genome. Comput. Chem. 26:195-206

Zhang M. 1997. Identification of protein coding regions in the human genome by quadratic discriminant analysis. Proc. Natl. Acad. Sci. 94:565–568

Zhang N., Xu Y., Akash M., McCouch S., and Oard J.H. 2005. Identification of candidate markers associated with agronomic traits in rice using Discriminant Analysis. Theor. Appl. Genet. 110:721-729

Zhuang J.Y, Fan Y.Y., Wu J.L., Xia Y.W., and Zheng K.L. 2000. Comparison of QTL detection for yield traits using F_2 and recombinant inbred lines in rice. Rice Genetics Newsletter 17:49-50

CHAPTER 2: IDENTIFICATION OF CANDIDATE MARKERS ASSOCIATED WITH AGRONOMIC TRAITS IN RICE USING DISCRIMINANT ANALYSIS[1, 2]

2.1 Introduction

Marker-assisted selection has been proposed as a complementary tool in plant improvement when reliable phenotyping and selection of complex traits is difficult or inefficient (Morgante and Salamini 2003; Xu et al. 2002). The initial task in this process typically requires screening potential parents for polymorphic molecular markers and the subsequent production of segregating or recombinant inbred populations. Loci or intervals are then defined on pre-existing genetic maps that are linked with a trait of interest by single-factor ANOVA (Jermstad et al. 2003), regression (Wang et al. 2004), interval (Lincoln et al. 1992), or other standard mapping procedures. For complex quantitative traits, ≥300 recombinant inbred lines are generally evaluated, which require 3 to 4 years to develop. Moreover, relatively few meiotic events in F_2 or recombinant inbred lines limit the power of linkage analysis to dissect traits governed by multiple loci, and examination of genetic diversity in diploids is restricted to only two alleles segregating per locus (Flint-Garcia et al. 2003). Production of large segregating or intermating populations can promote recombination, but substantial investments in time, labor, and financial resources over multiple generations are required.

Association or linkage disequilibrium (LD) mapping, based on pairwise comparisons between observed and expected haplotype frequencies, has been used extensively in human studies (Cardon and Abecasis 2003) and recently in maize among polymorphic pairs of SNPs, and insertions/deletions of individual candidate genes for maturity and plant height (Remington et al. 2001; Thornsberry et al. 2001). Garris et al. (2003) characterized LD in the candidate

[1] Published in Theor Appl Genet 110:721-729, 2005.
[2] Reprinted with kind permission of Springer Science and Business Media.

region of xa5, a recessive gene conferring race-specific resistance to bacterial blight disease in rice. Thirteen segments from a 70-kb candidate region in 114 landrace accessions were sequenced along with five additional segments from an adjacent 45-kb region in resistant accessions. The results showed significant LD up to 100 kb among sites that suggested genome-wide scanning may be feasible for markers that are associated with agronomic traits. The candidate gene approach was recently employed in LD mapping of QTLs for disease and maturity traits in tetraploid potato (Gebhardt et al. 2004; Simko et al. 2004).

In the study reported here, we evaluated the potential of DA, a multivariate statistical procedure first developed by Fisher (1936), to identify candidate markers associated with agronomic traits among inbred lines of rice. This method involves the creation of two "training samples" derived from, in this case, selected inbred lines with contrasting phenotypic values. From DNA profiles of all inbred lines included in the experiment, markers are identified by DA that best differentiate among the training samples. An error rate, referred to as "percent correct classification," is calculated to measure the ability of the markers to correctly assign individual lines to the training samples. With high levels of correct classification, an association between marker and phenotype or agronomic trait is inferred.

DA has been used in plant research for diversity analysis of wild emmer wheat and species of Aster (Cammareri et al. 2004; Fahima et al. 2002) identification of drought-tolerant Kentucky bluegrass cultivars using morphological criteria (Ebdon et al. 1998), and to estimate position and effects of QTLs in simulated full and half-sib families (Gilbert and Le Roy 2003). Microarray expression profiling studies have utilized DA to identify genes and gene clusters associated with human diseases (DePrimo et al. 2003; Kari et al. 2003; Mendez et al. 2002; Musumarra et al. 2003) and to detect protein-coding regions in genomic sequences (Zhang 1998;

Zhang et al. 2002). Finally, DA procedures were used recently to accurately assign unrelated sweet potato clones, using AFLP markers to groups defined by high and low dry matter content (Mcharo et al. 2004).

The objective of this research was to assess the ability of DA, coupled with other procedures described here, to identify candidate markers associated with 12 agronomic traits among US and Asian rice inbred lines. Different training samples were created for each trait, and the corresponding percent correct classification was determined. The potential genetic basis of the DA-selected markers was evaluated by comparing their map locations with QTL markers previously identified by traditional mapping approaches.

2.2 Materials and Methods

2.2.1 Plant Materials

A total of 123 US lines were randomly selected from California (34 lines), Texas (35 lines), Arkansas (28 lines), Louisiana (24 lines), Mississippi (1 line), and Missouri (1 line). In addition, 95 rice lines from 17 countries of Asian, African, and South American origin were included. For field studies, each rice line was transplanted into a plot each with four rows and 32 plants at Alvin, Texas, during the summer of 1996 and 1997. Eight plants from the center of each plot were evaluated to determine characteristic phenotypes, including plant height (ground to tip of tallest panicle), heading date (days from planting to 50% of plants flowered), tiller number, panicle length, 1,000-grain weight, grain length, grain width, grain length/width ratio, grain thickness, flag leaf length, flag leaf width, and stem diameter. One productive tiller of each selected plant was taken for measurement of stem diameter, flag leaf length, and width. From each line three typical plants were selected as “type specimens” for panicle harvesting. Three panicles from each of the three typical plants were then evaluated for panicle length. Ten seeds

from each line were used to measure grain length, width, and thickness. Two samples from each entry were used to obtain data for 1,000-grain weight. Data were averaged across each trait and line, and an ANOVA was carried out (PROC MIXED, SAS Institute, version 9.0) to detect differences among mean values of US and Asian lines. The type specimens were used as seed sources for molecular analysis.

2.2.2 Discriminant Analysis and Associated Procedures

DNA profiles were obtained for lines, using 60 SSR and 114 RFLP markers selected randomly over the 12 rice chromosomes at ~10- to 12-cM intervals (for additional details, see Xu et al. 2004).

To analyze phenotypic data, the following procedures were carried out:

1. Transformed data if necessary to normal distribution by log, square root, or other methods.
2. Used one, two, or three standard deviations of trait distribution to create user-defined training samples.

For molecular data analysis:

1. Transformed raw marker data to identify individual alleles.
2. Filled in missing marker data, using the Multiple Imputation procedure (SAS Institute ver. 9.0).
3. Performed molecular analysis of variance (AMOVA, Excoffier et al. 1992) of marker profiles to test differences among training samples using Arlequin software (Schneider et al. 2002).
4. Identified potential population structure by genetic distance (http://www.powermarker.net) or model-based (http://www.stats.ox.ac.uk/ ~pritch/home.html) method.

5. Performed parametric discriminant analysis (PROC STEPDISC, SAS Institute ver. 9.0, forward method, select up to 15 alleles, minimum criteria set with default SLENTRY=0.15) to identify marker(s) that best differentiate training samples within each subpopulation.
6. Used nonparametric method within the DISCRIM procedure (SAS Institute ver. 9.0) to perform k-nearest-neighbor classification of inbred lines into pre-defined groups.
7. Calculated percent correct classification with *cross-validate* option within the PROC DISCRIM procedure (SAS Institute ver. 9.0).

SSR and RFLP markers were located on the Rice–Cornell SSR 2001-1 and /or Rice–Cornell RFLP 2001–2002 genetic maps (http://www.gramene.org). Polymorphism information content (PIC) and gene diversity index (GDI) values were calculated using the PowerMarker program (http://www.powermarker.net). Linear correlations among traits were obtained using PROC CORR (SAS Institute ver. 9.0).

2.3 Results and Discussion

The US and Asian lines exhibited a wide range of phenotypic diversity for all 12 traits measured under Texas field-plot conditions (Table 2.1). Mean values for flag leaf width, panicle length, and 1,000-grain weight were not significantly different between US and Asian lines. The US material produced greater grain length and grain length/width ratio than the Asian germplasm, while plant height, heading date, flag leaf length, tiller number, stem diameter, grain width, and grain thickness showed greater mean values in Asian versus US lines. Heading date was weakly to moderately correlated with plant height, stem diameter, and flag leaf length (r=0.38, 0.31, and 0.36, respectively, P<0.001 for all). Plant height was moderately correlated with panicle length, stem diameter, and flag leaf length (r=0.58, 0.61, and 0.51, P<0.001 for all). While productive tiller number was not correlated with any character, 1,000-grain weight as a

Table 2.1. Mean values of agronomic traits of US and Asian rice lines, 1996-1997, Alvin, Texas.

Trait	US + Asian lines		US lines		Asian lines		US vs Asian P-value[a]
	Mean ± SD	Range	Mean ± SD	Range	Mean ± SD	Range	
Plant height (cm)	106.54 ± 20.64	63.00-185.40	101.76± 16.69	71.80-138.40	112.72± 23.54	63.00-185.40	0.001
Heading date (d)	95.79 ± 9.17	75.00-129.00	94.01± 7.00	82.00-119.00	98.11± 11.01	75.00-129.00	0.001
Flag leaf length (cm)	34.04 ± 6.03	20.40-53.60	33.13± 5.49	20.40-48.60	35.23± 6.51	23.20-53.60	0.011
Flag leaf width (cm)	1.49 ± 0.26	1.00-2.40	1.47± 0.21	1.10-2.10	1.50± 0.31	1.00-2.40	0.339
Tiller number	12.03 ± 8.33	4.20-50.20	7.76± 2.42	4.20-16.20	17.57± 9.89	5.40-50.20	0.001
Stem diameter (mm)	5.05 ± 0.91	2.90-8.30	4.93± 0.75	3.40-7.20	5.22± 1.07	2.90-8.30	0.021
Panicle length (cm)	22.87 ± 2.99	15.90-31.00	22.62± 2.93	15.90-29.20	23.21± 3.05	16.20-31.00	0.150
Grain length (L) (mm)	8.94 ± 1.05	6.80-12.20	9.14± 0.95	7.20-11.10	8.67± 1.12	6.80-12.20	0.001
Grain width (W) (mm)	2.95 ± 0.41	2.20-4.00	2.88± 0.42	2.20-4.00	3.05± 0.39	2.30-3.80	0.003
Grain L/W ratio	3.11 ± 0.70	1.89-4.83	3.27± 0.70	2.00-4.32	2.91± 0.64	1.89-4.83	0.001
Grain thickness (mm)	2.09 ± 0.15	1.80-2.50	2.07± 0.14	1.80-2.50	2.12 ± 0.15	1.80-2.40	0.017
1000 grain weight (g)	25.73 ± 3.53	14.30-36.70	25.35± 3.26	14.30-33.30	26.23± 3.82	17.20-36.70	0.070

[a] P values calculated from PROC MIXED, SAS Institute, ver. 9.0.

component of grain yield was associated with grain width and thickness (r=0.61 and 0.66, P<0.001 for both). The high level of trait diversity across the US and Asian germplasm was reflected in high levels of molecular variation, with 1,153 alleles detected across the 60 SSR and 114 RFLP sampled loci. When all lines were combined, the average number of alleles per locus, allele frequencies, and PIC values were similar to those previously reported by Xu et al. (2004). Greater diversity in the Asian versus US material was observed in all comparisons. For example, the mean PIC value for Asian lines was greater (0.499, range 0–0.913) than the US germplasm (0.269, range 0–0.882), and the mean number of Asian alleles per locus of 6.3 (range 1–29) was greater than the US accessions with a mean of 4.20 and a range of 1–25. The 12 monomorphic loci observed across all US and Asian lines (RZ386, CDO328, CDO524, RZ69, CDO244, RZ495, CDO89, RZ593, CDO544, CDO412, RZ900, and CDO1338) were produced using cDNA probes during the RFLP analysis. The markers CDO118, CDO395, CDO962, RZ2, RZ14, RZ87, RZ103A, RZ166, RZ499, RZ599, RZ783, and RZ836 were monomorphic in the Asian lines, but not in the US or combined material. The number of detected US monomorphic loci was the same as in the combined analysis plus six additional markers (BCD349, CDO36, CDO127A, CDO686, RZ141A, and RZ141B). Population structure analysis revealed three subpopulations, where subpopulation 1 consisted of 159 individuals, of which 136 (86%) were classified as *japonica* lines, and 117 (74%) were US lines. Subpopulations 2 and 3 were composed of 16 and 43 individuals, respectively, with 6/16 (37%) and 11/43 (25%) classified as *japonica* accessions and 3/16 (19%) and 3/43 (7%), respectively, were of US origin. The remaining individuals were classified as *indica* accessions. The range of phenotypic values overlapped for all traits among the three subpopulations (data not shown) as well as the US, Asian, and combined lines (Table 2.1). Mean values across subpopulations were similar for most traits except for the grain

measurements, where subpopulation 1 produced greater grain length, width, and weight values than those of subpopulation 2 and 3. GDI values of 0.29, 0.50, and 0.34 were observed for the three respective subpopulations, indicating that population 2, which is the smallest, is also the most diverse. A similar trend in mean PIC values for subpopulations 1, 2, and 3 showed moderate values of 0.26, 0.46, and 0.31, respectively, that bracketed the PIC value for the 236 US and Asian lines reported by Xu at al. (2004). Mean frequencies across all alleles within each subpopulation were nearly identical with each other (0.22, 0.24, and 0.26), but when the US material was compared with the Asian material, the mean frequency of alleles in the combined US group (0.24) was larger than in the Asian (0.16) or in the combined dataset (0.15). Overall, these results suggest that the extent of phenotypic and molecular diversity of each of the subpopulations was comparable to all lines combined, but that the Asian material was considerably more diverse that the US accessions.

Table 2.2 shows that the DA procedure correctly classified the rice lines into early or late heading groups, using 5–10 markers. For the remaining traits, 86–100% correct classification was obtained with 1–15 markers, using the 1 standard deviation (SD), 2SD, or 3SD training samples. Accuracy did increase in all cases, with increasing numbers of markers used within each defined group or training sample. Population structure appeared to have little impact on accuracy of classification for heading date (Table 2.2), most likely due to similar characteristics among subpopulations and combined lines discussed above and to a relatively large proportion of individuals (73%) occurring within subpopulation 1 used for analysis. The same trend was observed for the remaining traits.

Data from Table 2.3 and Fig. 2.1 show that DA-selected markers were detected in the same or nearby regions for agronomic traits as previously identified QTLs. For example, in a

Table 2.2. Percent correct classification of 218 inbred rice lines assigned to defined groups of early and late heading date as training samples using a k-nearest neighbor algorithm with and without consideration of population structure, 1996-1997, Alvin, Texas.

Training samples for selection	No. lines	Early / late heading groups	No. of DA-selected alleles and % correct classification			
			1	5	10	15
Assuming	218	1SD[a]	66	86	79	95
no structure		2SD	73	93	98	100
		3SD	84	99	100	ND[b]
Assuming structure						
Sub population 1[c]	159	1SD	71	86	93	94
		2SD	78	94	100	ND
		3SD	88	100	ND	ND

[a] 1SD = one standard deviation between early and late groups; 2SD = two standard deviations; 3SD = three standard deviations.
[b] No data obtained because all discriminating markers were selected.
[c] Only sub population 1 was evaluated because remaining two subpopulations contained insufficient size (n = 16, 43) for analysis.

comparison of 23 previous papers reporting the position of rice QTLs for similar traits, DA-selected allele RM263_156_2 mapped within the 7.2-cM QTL *qHD-2* (Zou et al. 2000) on chromosome 2 for heading date, and RM250_170_11 mapped within the 22.1-cM unnamed QTL reported by Brondani et al. (2002), near the bottom of chromosome 2. Three DA alleles, RM204_120_9, RM204_166_26, and RM204_104_2, were found on chromosome 6 for heading date within the 12-cM QTL *Hd6c* (Xing et al. 2001), the 31.4-cM *hd6* (Yu et al. 2002), and the 20.4-cM *dtm6.1* (Xiao et al. 1998). DA alleles RM248_102_12 and RM248_84_5 were detected within the 14.1-cM QTL *DTF1* (Brondana et al. 2002) for heading date, located on the bottom half of chromosome 7. The same two DA alleles were also found associated with this trait within the 12.9-cM QTL *Hd2* (Ishimaru et al. 2001; Lin et al. 1998; Yamamoto et al. 2000) and within the 59.4-cM interval *Qhd7* (Mei et al. 2003).

Table 2.3. SSR/RFLP alleles identified by DA from 1 SD, 2 SD and 3 SD training samples among 218 US and Asian inbred rice lines.

Trait	DA-selected RFLP/SSR alleles producing 96-100% correct classification		
	1SD Training Sample	2SD Training Sample	3SD Training Sample
Plant Height (cm)	CDO405_140_3[a], CDO580_6_1, RZ424B_26_1, RM212_136_7, RM263_156_2, RM247_154_12, RM21_132_2, RM235_96_2, RM1_89_6, RG716_154_4, RM232_156_10, RM22_187_3, RZ53_22_1, RM232_150_7, RM205_155_12	RM255_151_7, RM212_136_7, RM20B_207_1, RZ53_22_1, RM259_158_7, RM38_238_3, RM241_130_8, RM38_266_17, RM22_185_2, RM10_164_2	RM18_159_4, RM235_136_15, RM248_80_3, RZ404_97_5
Heading date (d)	RM263_156_2, RM248_92_8, RM255_149_6, RZ740_60_5, RG716_96_3, RM219_210_9, RG146_31_1, RM204_178_30, RM255_141_2, RM44_130_13, RM224_140_7, RM55_231_3, RM22_193_6, RZ53_22_1, RZ537A_64_8	RG716_96_3, RM248_102_12, RM263_156_2, RM247_154_12, RM204_120_9, RM84_110_3, RM255_153_8, RM202_180_10, RM204_166_26, RM250_170_11, RZ574_102_1, RM209_159_15	RM10_163_1, CDO78_63_2, RG716_96_3, RM10_179_12, RM248_84_5, RM11_125_2, RZ206B_69_2, RM204_104_2, CDO78_68_3, RM19_246_7, RM19_225_5, RM204_180_31
Flag leaf length (cm)	CDO456_36_6, RZ424B_26_1, RM235_136_15, RM19_222_4, RM263_160_4, RM230_259_6, RM257_173_24, RM204_144_18, CDO202_250_4, RM223_161_10, RZ424A_11_1, RM228_152_19, RG716_154_4, RM224_132_2	RM205_155_12, RM259_171_18, RZ424A_36_2, RM202_178_9, RM204_144_18, RM240_132_5, RM240_136_7, RM255_147_5, RZ405_43_1, RM20B_216_3	CDO456_36_6, RM224_134_3, RM224_156_11, RG1109_36_3
Flag leaf width (cm)	RM44_130_13, CDO78_63_2, RM44_120_12, RM215_148_3, RZ599_38_2, RM44_112_8, RZ53_22_1, RM233B_142_5, RM239_144_3, RM38_250_9, RM255_143_3, BCD98_44_3, RM14_189_8, RM14_171_2, RM10_179_12	RM11_127_3, RZ599_38_2, RM215_148_3, RM23_136_2, RZ588_190_3, RM209_163_18, RM14_209_15, RM19_249_8, RZ740_55_3, RM209_127_4, RM219_222_14, RM226_197_3, RM250_170_11, RM207_137_15	RM232_162_13, RM21_154_9, CDO545B_48_1, RM20A_285_13
Tiller number	RZ400_15_2[A], RM212_116_4, RM14_187_7, RM1_95_9, RZ404_33_2, RM259_162_11, RM204_140_16, RM226_193_1, RM13_151_10, RM240_132_5, RZ599_78_5, RM48_211_4, RZ141A_70_3, RM257_170_21	RM262_141_1, RM204_168_27	RM262_141_1
Stem diameter (mm)	RM255_151_7, RZ424A_11_1, RM222_213_9, RZ284_67_1, RM19_216_2, RM235_136_15, RM222_209_7, CDO405_180_5, RM21_164_14, RZ405_160_9, CDO98_50_1, RM27_158_4, RZ143_91_2, RZ103A_54_3	RM232_158_11, RZ53_200_5, RM262_157_6, CDO718_39_1, RM259_171_18, RM21_162_13, RM22_187_3, RM204_148_20, RZ53_22_1, RM257_177_27, RM222_201_3, RM259_158_7	RM228_114_5, RM259_156_5, RM263_156_2, RM232_156_10, RM44_130_13
Panicle length (cm)	RM7_175_7, RM232_144_4, RZ424B_26_1, RM224_134_3, RM20A_276_10, RM263_160_4, RM212_112_2, RZ141B_240_2, CDO405_170_4, RM38_266_17, RG716_86_2, RM235_96_2	RM14_183_6, RM55_235_5, RM207_117_6, RZ103A_46_1, RM7_175_7, RM20A_302_19, RM228_150_18, RM209_145_9, RM38_266_17, RM219_202_5	CDO405_170_4, RM224_157_12, RM207_125_10
Grain length (L) (mm)	RZ574_215_2, RM16_184_6, RG757_150_2, RM19_237_6, RM20B_207_1, RM1_93_8, RM228_120_8, RM1_117_16, RM205_127_6, RM239_144_3, RM257_185_31, RM14_187_7, RZ537A_26_1, RM219_212_10, Z599_78_5	RZ574_215_2, RM11_127_3, RZ405_158_8, RM18_151_2, RM253_140_15, RZ405_77_5, RM19_225_5, RM257_177_27	RM11_127_3
Grain width (W) (mm)	RM248_82_4, RM202_159_4, RM14_183_6, RM51_132_2, RM263_184_17, RZ625_180_4, RM232_164_14, RM247_162_16, RM247_172_20, RZ599_40_3, RZ400_32_3, RM207_117_6, RM10_175_10, RZ783_40_2	CDO365_160_4, RM14_183_6, RM232_160_12, RM55_219_1, CDO405_89_2, RM1_95_9, RM202_157_2, RM13_133_2	RM14_183_6, RM11_123_1
Grain L/W ratio	RZ574_215_2, RM258_150_7, RM21_162_13, RM21_154_9, RM202_184_12, RM209_161_16, RZ405_158_8, RM240_132_5, RM10_166_4, RM226_269_17, RM262_143_2, RM13_151_10, RZ599_78_5, RM226_219_9	RM14_183_6, RM21_160_12, RZ405_58_2, RM222_219_12, RM248_86_6, RM263_154_1, RM204_142_17, RM20A_269_7	RZ400_32_3
Grain thickness (mm)	ND[B]	RM14_183_6, RM18_161_5, RM14_197_12, RM257_177_27, RM224_157_12, RM259_162_11, RM209_127_4, RG322A_25_2, RM204_176_29, RM232_152_8	RZ329_33_2
1000 grain weight (g)	RM248_82_4, RM205_161_15, RG901_144_4, RM223_147_3, RM204_178_30, RM205_153_11, RM224_138_6, RM259_159_8, RM38_266_17, RM226_221_10, RM205_127_6, RM226_273_18	RM7_175_7, RM255_147_5, RZ424B_54_2, RM16_184_6, RZ599_38_2, RM215_156_7, RM202_159_4, RZ206B_69_2, CDO456_28_3, RM21_152_8	RZ329_43_3, RM44_92_2, CD0118_69_1, RM241_138_13

[a] First component of allele designation is SSR/RFLP marker, second is allele size in bp (SSR) or 100 bp (RFLP), third is allele number at locus. Allele order in table corresponds to its relative contribution to calculated discriminant rule.

[b] No alleles identified most likely due to lack of phenotypic variation observed among lines.

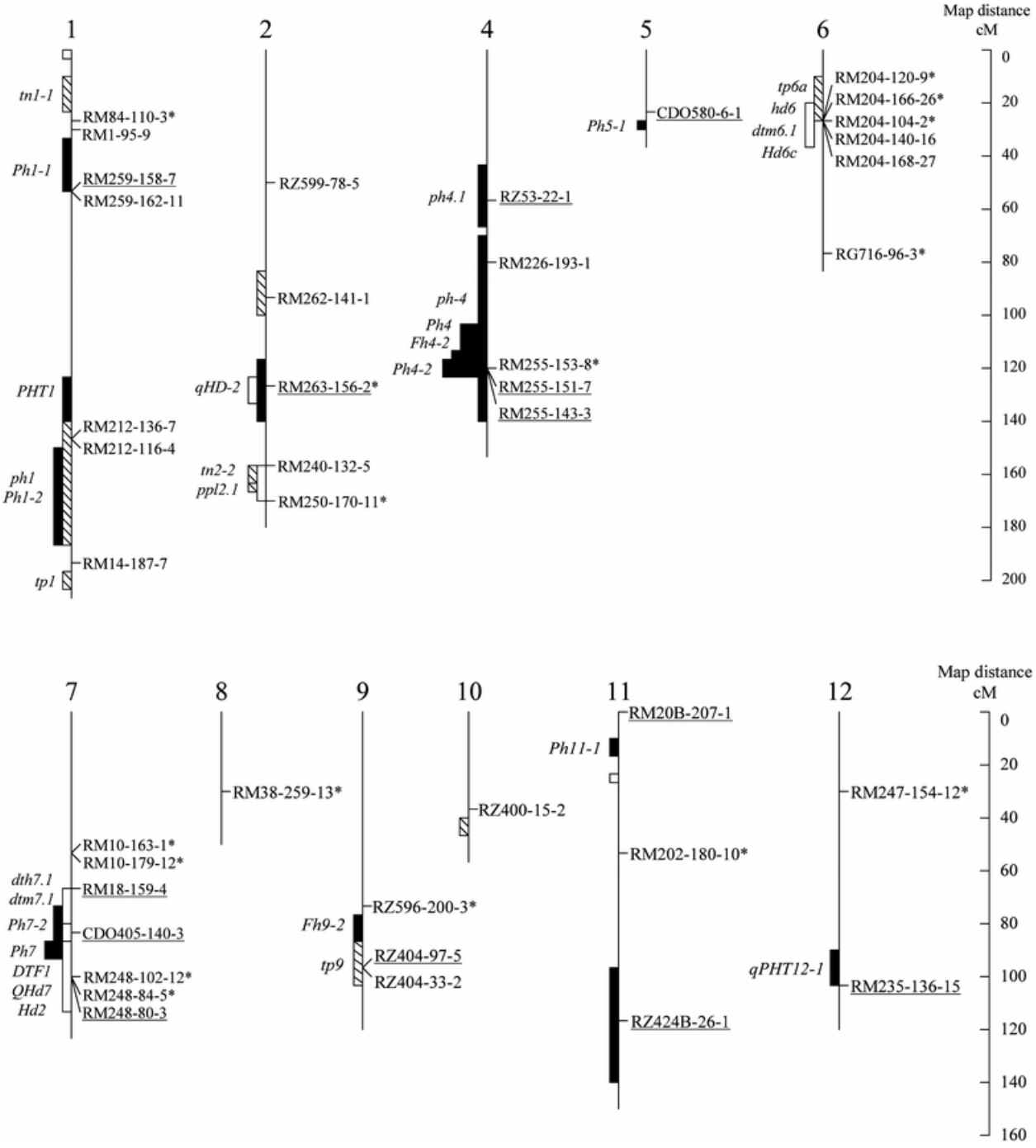

Fig. 2.1. Chromosomal locations of markers identified by discriminant analysis (DA) and quantitative trait loci (QTL) mapping for plant height, maturity, and tiller number in rice. *Solid*, *empty*, and *striped boxes* represent QTLs detected in previous research using standard IM/CIM methods for plant height, heading date, and tiller number, respectively. DA-selected SSR or RFLP markers associated with plant height are *underlined*. DA-selected SSR or RFLP markers associated with heading date are labeled with an *asterisk*. DA-selected SSR or RFLP markers with no label are associated with tiller number.

For plant height, DA allele RM259_158_7 mapped to chromosome 1 within the 15.9-cM QTL *Ph1-1* reported by Cao et al. 2001a (Fig. 2.1). DA-selected allele RM212_136_7 was located 4.3 cM from the 24.3-cM QTL *PHT1* (Brondani et al. 2002), 4.8 cM from the 33-cM *ph1* (Yu et al. 2002), and 7.3 cM from the 48.6-cM *Ph1-2* (Cao et al. 2001a), all on the bottom half of chromosome 1. The RM263_156_2 allele selected by DA mapped within the 24-cM unnamed QTL for plant height reported by Mei et al. (2003) on chromosome 2. Allele RZ53_22_1 was found within the 15.5-cM interval *ph4.1* (Moncada et al. 2001) on chromosome 4. RM255_151_7 and RM255_143_3 mapped within four overlapping intervals on chromosome 4 for plant height: the 2.7-cM *Ph4-2* (Cao et al. 2001a), the 42-cM *ph4* (Yan et al.1999), the 58.6-cM *ph-4* (Lu et al. 1997), the 42 cM *Fh4-2* (Cao et al. 2001b), and within the 11.8-cM unnamed QTL reported by Fang and Wu (2001). The CDO580-6-1 allele was detected 1.3 cM from the 6.6-cM QTL *Ph5-1* (Yan et al. 1998a) located at the bottom of chromosome 5. DA allele CDO405_140_3 mapped within the 9.7-cM interval *Ph7-2* (Cao et al. 2001a), 5 cM from an unnamed 7.6-cM QTL reported by Ishimaru et al. (2001), and 6.8 cM from the 5.8-cM *Ph7* (Yan et al. 1998a), all on chromosome 7. Allele RZ404_97_5 was detected 7.5 cM from the 0.16-cM *Fh9-2* (Cao et al. 2001b) on chromosome 9. DA allele RZ424B_26_1 mapped within an 18.5-cM unnamed QTL (Mei et al. 2003) on chromosome 11. DA allele RM235_136_15 was found within the 6.9-cM *qPHT12-1* (Hemamalini et al. 2000), near the bottom of chromosome 12.

For tiller number, RM1_95_9 mapped 5 cM from the 14.1-cM *tn1-1* (Yan et al. 1998b), near the top of chromosome 1, and RM14_187_7 was found 3.5 cM from the 8.3-cM *tp1* (Hua et al. 2002) at the bottom of the same chromosome (Fig. 2.1). Allele RM212_116_4 was found within a 45.5-cM unnamed QTL reported by Lafitte et al. (2002) on chromosome 1, and allele RM262_141_1 mapped within the 22.9-cM unnamed QTL detected by Shen et al. (2001) on

chromosome 2. RM240_132_5 was observed within the 4.9-cM *tn2-2* (Yan et al. 1998b), the 0.6-cM *ppl2.1* (Xiao et al. 1998), and within an 8-cM unnamed QTL reported by Lafitte et al. (2002) at the bottom of chromosome 2. Alleles RM204_140_16 and RM204_168_27 were detected within the 25-cM QTL *tp6a* (Hua et al. 2003) on chromosome 6. DA-selected allele RZ404_33_2 was found within the 12.9-cM *tp9* (Hua et al. 2003) and within an 11.8-cM unnamed QTL (Liao et al. 2001), near the bottom of chromosome 9. Finally, DA allele RZ400 _15_2 mapped 8.3 cM from an 8-cM unnamed QTL reported by Liao et al. (2001), near the bottom of chromosome 10. Figure 2.1 shows seven loci (RM259, RM263, RM212, RM255, RM204, RM248, and RZ404) on chromosomes 1, 2, 4, 6, 7, and 9 associated with more than one trait, which suggests that these markers may be associated with pleotropic or closely linked genes for the corresponding characters. DA-selected markers associated with the remaining nine traits also mapped within or nearby previously reported QTLs (data not shown).

In addition to the DA alleles that pointed to the same or nearby regions as previously reported QTLs, Fig. 2.1 shows several DA-selected markers not found by traditional methods. For example, the following alleles selected for heading date were not found associated with previously reported QTLs: RM84_110_3 on chromosome 1, RM255_153_8 on chromosome 4, RG716_96_3 on chromosome 6, RM10_163_1 and RM10_179_12 on chromosome 7, RM38_259_13 on chromosome 8, RZ596 _200_3 on chromosome 9, RM202_180_10 on chromosome11, and RM247_154_12 on chromosome 12. Similarly, DA alleles not associated with reported QTLs for plant height include RM18_159_4 and RM248_80_3 on chromosome 7 and RM20B_207_1 on chromosome 11. For tiller number, DA-selected alleles RM259_162_11 on chromosome 1, RZ599_78_5 on chromosome 2, and RM226_193_1 on chromosome 4 were found at positions other than the corresponding QTLs reported in the literature. These markers

identified by DA may therefore represent new loci associated with plant height, maturity, and vigor. Similar results were obtained for the remaining nine agronomic traits evaluated in this study (data not shown).

2.4 Conclusions

Results from this study indicate that marker alleles associated with all traits were identified by DA among inbred rice lines at high levels of correct percent classification within subpopulations and across all lines. Cross-validation results and a comparison of DA- and QTL-selected markers on the rice genetic map suggest that this approach can efficiently identify markers from multiple germplasm sources. The DA statistical model is built upon various assumptions, including normality of data and homogeneity of covariance matrices that appear to be poorly satisfied by SSR/RFLP marker data in this study. However, Lachenbruch (1975) and Klecka (1980) point out that even with modest violations of these assumptions, DA is relatively robust when using categorical data such as the molecular profiles from this study. Therefore, our conclusions of marker–trait associations based on DA analysis should not be adversely affected, which is supported by our DA–QTL genetic map comparisons.

Relatively high levels of molecular and phenotypic diversity of the US and Asian lines, compared with typical progeny from a single cross, most likely contributed to the ability of DA to identify putative alleles associated with the agronomic characters. Population structure appeared to have minimal impact on the ability of DA-selected markers to correctly assign individuals in this study to predefined phenotypic groups or to map to regions identified in previous QTL experiments. However, population structure has been shown to have a dramatic effect on DA analysis of other rice populations (Aluko and Oard, unpublished results), so this step should always be included as part of the DA procedure described here. Because the level of

linkage disequilibrium, i.e., the nonrandom association of loci, can be affected by breeding history, additional DA studies of this issue will be required.

The potential advantages of the DA approach reported here include the ability to simultaneously evaluate numerous loci with multiple alleles across a wide range of inbred lines for association with simple or complex agronomic traits. Additional genetic analysis of the DA-selected markers in segregating populations derived from controlled crosses will be required to confirm the putative association of the alleles identified in this research with the agronomic traits.

2.5 References

Brondani C., Rangel N., Brondani V., and Ferreira E. 2002. QTL mapping and introgression of yield-related traits from Oryza glumaepatula to cultivated rice (*Oryza sativa*) using microsatellite markers. Theor. Appl. Genet. 104:1192-1203

Cammareri M., Errico A., Sebastiano A., and Conicella C. 2004. Genetic relationships among *Aster* species by multivariate analysis and AFLP markers. Hereditas 140:193-200

Cao G., Zhu J., He C., Gao Y., Yan J., and Wu P. 2001a. Impact of epistasis and QTL x environment interaction on the developmental behavior of plant height in rice (*Oryza sativa* L.). Theor. Appl. Genet. 103:153-160

Cao G.Q., Zhu J., He C.X., Gao Y.M., and Wu P. 2001b. QTL analysis for epistatic effects and QTL x environment interaction effects on final height of rice (*Oryza sativa* L.) Acta Genetica Sinica 28:135-143

Cardon L.R. and Abecasis G.R. 2003. Using haplotype blocks to map human complex trait loci. Trends Genet. 19:135-40

DePrimo S.E., Wong L.M., Khatry D.B., Nicholas S.L., Manning W.C., Smolich B.D., O'Farrell A.M., and Cherrington J.M. 2003. Expression profiling of blood samples from an SU5416 Phase III metastatic colorectal cancer clinical trial: a novel strategy for biomarker identification. BMC Cancer 3:3

Ebdon J.S., Petrovic A.M., and Schwager S.J. 1998. Evaluation of Discriminant Analysis in identification of low and high-water use Kentucky bluegrass cultivars. Crop Sci. 38:152-157

Excoffier L., Smouse P.E., and Quattro J.M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479-491

Fahima T., Roder M.S., Wendehake K., Kirzhner V.M., and Nevo E. 2002. Microsatellite polymorphism in natural populations of wild emmer wheat, *Triticum dicoccoides*, in Israel. Theor. Appl. Genet. 104:17-29

Fang P. and Wu P. 2001. QTL x N-level interaction for plant height in rice (*Oryza sativa* L.) Plant and Soil 236:237-242

Fisher R.A. 1936. The use of multiple measurements in taxonomic problems. Ann Eugenics 7:179-188

Flint-Garcia S.A., Thornsberry J.M., and Buckler E.S. 2003. Structure of linkage disequilibrium in plants. Ann. Rev. Plant Biol. 54:357-374

Garris A.J., McCouch S.R., and Kresovich S. 2003. Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the xa5 locus of rice (*Oryza sativa* L.). Genetics 165:759-769

Gebhardt C., Ballvora A., Walkemeier B., Oberhagemann P., and Schuler K. 2004. Assessing genetic potential in germplasm collections of crop plants by marker-trait association: a case study for potatoes with quantitative variation of resistance to late blight and maturity type. Mol. Breed. 13:93-102

Gilbert H. and Le Roy P. 2003. Comparison of three multitrait methods for QTL detection. Genet. Sel. Evol. 35:281-304

Hemamalini G.S., Shashidhar H.E., and Hittalmani S. 2000. Molecular marker assisted tagging of morphological and physiological traits under two contrasting moisture regimes at peak vegetative stage in rice (*Oryza sativa* L.). Euphytica 112:69-78

Hua J., Xing Y., Wu W., Xu C., Sun X., Yu S., and Zhang Q. 2003. Single-locus heterotic effects and dominance by dominance interactions can adequately explain the genetic basis of heterosis in an elite rice hybrid. PNAS 100:2574-2579

Hua J.P., Xing Y.Z., Xu C.G., Sun X.L., Yu S.B., and Zhang Q. (2002) Genetic dissection of an elite rice hybrid revealed that heterozygotes are not always advantageous for performance. Genetics 162: 1885-1895

Ishimaru K., Yano M., Aoki N., Ono K., Hirose T., Lin S.Y., Monna L., Sasaki T., and Ohsugi R. 2001. Toward the mapping of physiological and agronomic characters on a rice function map: QTL analysis and comparison between QTLs and expressed sequence tags. Theor. Appl. Genet. 102:793-800

Jermstad K.D., Bassoni D.L., Jech K.S., Ritchie G.A., Wheeler N.C., and Neale D.B. 2003. Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas fir. III. Quantitative trait loci-by-environment interactions. Genetics 165:1489-506

Kari L., Loboda A., Nebozhyn M., Rook A.H., Vonderheid E.C., Nichols C., Virok D., Chang C., Horng W.-H., Johnston J., Wysocka M., Showe M.K., and Shwe L.C. 2003. Classification and prediction of survival in patients with the leukemic phase of cutaneous T cell lymphoma. J. Exp. Med. 197:1477-1488

Klecka W.R. 1980. Discriminant analysis. Quantitative applications in the social sciences series, no.19. Sage, Thousand Oaks

Lachenbruch P.A. 1975. Discriminant analysis. Hafner, New York

Lafitte H.R., Courtois B., and Arraudeau M. 2002. Genetic improvement of rice in aerobic systems: Progress from yield to genes. Field Crops Research 75:171-190

Liao C.Y., Wu P., Hu B., Yi K.K. 2001. Effects of genetic background and environment on QTLs and epistasis for rice (*Oryza sativa* L.) panicle number. Theor. Appl. Genet. 103:104-111

Lin S.Y., Sasaki T., and Yano M. 1998. Mapping quantitative trait loci controlling seed dormancy and heading date in rice, *Oryza sativa* L., using backcross inbred lines. Theor. Appl. Genet. 96:997-1003

Lincoln S., Daly M., and Lander E. 1992. Constructing genetic linkage maps with MAPMAKER/EXP. Whitehead Institute Technical Report. Available via http://www-genome.wi.mit.edu/ ftp/distribution/software

Lu C.F., Shen L.H., Tan Z.B., Xu Y.B., He P., Chen Y., and Zhu L.H. 1997. Comparative mapping of QTLs for agronomic traits of rice across environments by using a doubled-haploid population. Theor. Appl. Genet. 94: 145–150

Mcharo M., Labonte D.R., Oard J.H., Kays S.J., and McLaurin W.J.. 2004. Linking quantitative traits with AFLP markers in sweetpotato using Discriminant Analysis. Acta Horticultura 637:285-293

Mei H.W., Luo L.J., Ying C.S., Wang Y.P., Yu X.Q., Guo L.B., Paterson A.H., and Li Z.K. 2003. Gene actions of QTLs affecting several agronomic traits resolved in a recombinant inbred rice population and two testcross populations. Theor. Appl. Genet. 107:89-101

Mendez M.A., Hodar C., Vulpe C., Gonzalez M., and Cambiazo V. 2002. Discriminant analysis to evaluate clustering of gene expression data. FEBS Letters 522:24-28.

Moncada P., Martínez C.P., Borrero J., Chatel M., Gauch Jr H., Guimaraes E., Tohme J., and McCouch S.R. 2001. Quantitative trait loci for yield and yield components in an *Oryza sativa* × *Oryza rufipogon* BC2F2 population evaluated in an upland environment. Theor. Appl. Genet. 102:41–52

Morgante M. and Salamini F. 2003. From plant genomics to breeding practice. Curr. Opin. Biotechnol. 14:214-9

Musumarra G., Barresi V., Condorelli D.F., and Scirè S. 2003. A bioinformatic approach to the identification of candidate genes for the development of new cancer diagnostics. Biological Chemistry 384:321-327

Remington D.L., Thornsberry J.M., Matsuoka Y., Wilson L.M., Whitt S.R., Doebley J., Kresovich S., Goodman M.M., and Buckler E.S. 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc. Natl. Acad. Sci. 98:1147-1184

Schneider S., Roessli D., and Excoffier L. 2002. Arlequin ver 2.000: A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva,Seitzerland

Shen L., Courtois B., McNally K.L., Robin S., and Li Z. 2001. Evaluation of near-isogenic lines of rice introgressed with QTLs for root depth through marker-aided selection. Theor. Appl. Genet. 103:75-83

Simko I., Costanzo S., Haynes K.G., Christ B.J., and Jones R.W. 2004. Linkage disequilibrium mapping of a Verticillium dahliae resistance quantitative trait locus in tetraploid potato (*Solanum tuberosum*) through a candidate gene approach. Theor. Appl. Genet. 108:217-224

Thornsberry J.M., Goodman M.M., Doebley J., Kresovich S., Nielsen D., and Buckler E.S. 2001. *Dwarf8* polymorphisms associate with variation in flowering time. Nature Genetics 28:286-289

Wang D., Graef G.L., Procopiuk A.M., and Diers B.W. 2004. Identification of putative QTL that underlie yield in interspecific soybean backcross populations. Theor. App. Genet. 108:458-467

Xiao J.H., Li J., Grandillo S., Sang N.A., Yuan L., Tanksley S.D., and McCouch S.R. 1998. Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. Genetics 150: 899-909

Xing Y.Z., Xu C.G., Hua J.P., Tan Y.F., and Sun X.L. 2001. Mapping and isolation of quantitative trait loci controlling plant height and heading date in rice. Acta Bot. Sinica 43:721-726

Xu Y. 2002. Global view of QTL: Rice as a model. In M.S. Kang (ed), Quantitative Genetics, Genomics and plant Breeding. CAB International. pp 109-134

Xu Y., Beachell H., and McCouch S.R. 2004. A marker-based approach to broadening the genetic baseof rice (*Oryza sativa* L.) in the USA. Crop Sci. 14:1947-1959

Yamamoto T., Lin H., Sasaki T., and Yano M. 2000. Identification of heading date quantitative trait locus *Hd6* and characterization of its epistatic interactions with *Hd2* in rice using advanced backcross progeny. Genetics 154:885-891

Yan J., Zhu J., He C., Benmoussa M., and Wu P. 1999. Molecular marker-assisted dissection of genotype x environment interaction for plant. Crop Sci. 39: 538-544

Yan J.Q., Zhu J., He C., Benmoussa M., and Wu P. 1998a. Molecular dissection of developmental behavior of plant height in rice (*Oryza sativa* L.). Genetics 150:1257-1265

Yan J.Q., Zhu J., He C.X., Benmoussa M., and Wu P. 1998b. Quantitative trait loci analysis for the developmental behavior of tiller number in rice (*Oryza sativa* L.) Theor. Appl. Genet. l.97:267-274

Yu S.B., Li J.X., Xu C.G., Tan Y.F., Li X.H., and Zhang Q. 2002. Identification of quantitative trait loci and epistatic interactions for plant height and heading date in rice. Theor. Appl. Genet. 104:619-625

Zhang C.T., Wang J., and Zhang R. 2002. Using a Euclid distance discriminant method to find protein coding genes in the yeast genome. Comput. Chem. 26:195-206

Zhang M.Q. 1998. Identification of protein-coding regions in Arabidopsis thaliana genome based on quadric discriminant analysis. Plant Mol. Biol. 37:803-806

Zou J.H., Pan X.B., Chen Z.X., Xu J.Y., Lu J.F., Zhai W.X., and Zhu L.H. 2000. Mapping quantitative trait loci controlling sheath blight resistance in two rice cultivars (*Oryza sativa* L.) Theor. Appl. Genet. 101:569–573

CHAPTER 3: STEPWISE, LEAST ANGLE AND LASSO REGRESSION MODELS FOR MARKER-TRAIT ASSOCIATIONS AMONG INBRED LINES OF RICE

3.1 Introduction

Association genetics has become increasingly popular as a means to rapidly and efficiently identify molecular markers associated with simple and complex traits. However, there are currently no clear or concrete methods to accurately select robust markers for all populations exhibiting different breeding or evolutionary histories. Reasons for this include population structure, admixture, and other factors that can lead to false positive associations and erroneous conclusions. We previously evaluated discriminant analysis (DA) to detect candidate RFLP and microsatellite markers associated with economically important traits in a large population of unrelated U.S. and Asian inbred lines of rice (Zhang et al., 2005). Associated marker alleles detected by DA mapped within the same genetic intervals when compared with previous traditional quantitative trait loci (QTL) mapping experiments that evaluated progeny from controlled crosses.

Multiple regression (MR) analysis had been used to identify markers associated with specific traits in many research projects. For example, Virk et al. (1996) used MR to identify associations between various quantitative traits and RAPD molecular markers with diverse Asian rice germplasm. In diploid alfalfa, three QTLs with highly significant effects on multinucleate-microspore formation were identified by ANOVA and stepwise MR (Tavoletti et al. 2000). In the analysis of genetic resources and adaptation in *Phytolacca dodecandra* L'Hér., 17 Ethiopian populations (249 individuals) were sampled along altitudinal gradients that varied from 1600 to 3000 m and MR showed a strong association between some RAPD markers, altitude, temperature and rainfall (Semagn et al. 2000). Stepwise MR was used in two maize inbred line crosses to identify nine yield QTLs, five of which were in the same regions as those identified by

composite interval mapping (Kraja and Dudley 2000). Similar results were found for plant and ear height, but for grain moisture, composite interval mapping identified nearly twice as many QTLs as stepwise MR (Kraja and Dudley 2000). MR of early yield on eight yield-related traits in cassava revealed harvest index, dry foliage weight and root diameter as the most important factors associated with early yield. Furthermore, based on single-marker regression analysis, QTLs were detected for early yield and associated traits (Okogbenin and Fregene 2002). Kraakman et al. (2004) employed step-wise multiple linear regression to find markers associated with grain yield and stability in barley.

Epistasis refers to the phenotypic effects of interactions among alleles at multiple loci. Our current understanding of biochemical and physiological genetics, as well as the regulation of gene expression, strongly suggests the ubiquitous nature of interactions among gene products. Morphological markers were used to demonstrate the existence of digenic epistatic interactions in barley populations long before the availability of any molecular tools (Fasoulas and Allard 1962). Recent genetic analyses using molecular markers in several plant species have clearly shown that, in addition to single locus QTLs, epistatic interactions play an important role in the genetic basis of quantitative traits (Lark et al. 1995). Li et al. (1997) found that epistasis is an important factor for complex traits such as yield components, especially those with low heritability such as grain number per panicle and grain weight per panicle. Zhuang et al. (2002) analyzed QTLs conditioning grain yield and five yield component traits at the one-locus and two-locus levels by using an RIL population derived from an indica-indica cross Zhenshan 97B × Milyang 46. Thirty-one QTLs detected showed significant additive effects for yield traits, of which 12 also exhibited significant epistatic effects. Xing et al. (2002) used a rice RIL population from the Zhenshan 97 × Minghui 63 cross for four yield and yield-component traits

that detected 29 QTLs of main effects, and 35 digenic interactions involving 58 loci. Epistatic interactions of panicle traits were also reported by Cui et al. (2002). Storey et al. (2005) used stepwise forward regression to sequentially select two significant markers and allowed for epistatic interactions for each gene expression trait in yeast. The results showed that epistatic interactions contribute to gene expression variation for at least 14% of all traits.

In the study reported here, we used the newly released SAS GLMSELECT procedure to select both additive and epistatic models for comparison with discriminant analysis (DA). A variety of model selection methods were available in this procedure, including the least absolute shrinkage and selection operator (LASSO) method of Tibshirani (1996) and the related least angle regression (LAR) method of Efron et al. (2004) designed to increase power and precision. The stepwise forward selection technique begins with just the intercept of the linear regression model and then sequentially adds the effects that most improve the fit. The process terminates when no significant improvement can be obtained by adding any effect (Documentation of The GLMSELECT Procedure (Experimental), SAS Institute 2005). LAR not only provides a selection method in its own right, but with one additional modification it can be used to efficiently produce LASSO solutions. The algorithm starts with all coefficients equal to zero, then finds the effect most correlated with the response, increases the coefficient in the direction of the sign of its correlation with the response, takes residuals along the way, and stops when some other predictor has as much correlation with the residual as it has, increases the first two coefficients in their joint least squares direction until some other predictor has as much correlation with the residual, and continues until all predictors are in the model (Documentation of The GLMSELECT Procedure (Experimental), SAS Institute 2005). The LASSO is a shrinkage and selection method for linear regression. It minimizes the usual sum of squared

errors, with a limit on the sum of the absolute values of the coefficients. When the limit is large enough, the constraint has no effect and the solution is just the usual multiple linear least squares regression. However, with smaller limit values (>=0), the solutions are shrunken versions of the least squares estimates. Often, some of the variable coefficients are zeros that reduce the number of variables in the model. Choosing the limit is similar to choosing the number of predictors to use in a regression model, and cross-validation is a good tool for estimating the best value for the limit (www-stat.stanford.edu/~tibs/lasso/simple.html).

The principal objective of this research was to compare different regression models, with and without interaction terms and population structure for identifying marker alleles associated with quantitative traits in a rice population composed of U.S. and Asian inbred accessions.

3.2 Materials and Methods

A total of 218 inbred lines from U.S. and Asia were grown in single-row plots in 1996 and 1997 near Alvin, Texas as previously described by Zhang et al. 2005. Traits measured: plant height, heading date, tiller number, panicle length, grain weight, length, width, length-width ratio, thickness, flag leaf length, width, and stem diameter. DNA profiles obtained for lines using 60 SSR and 114 RFLPs were selected randomly over the genome. For phenotypic data: (1) Transform data if necessary to normal distribution by log, square root or other method; (2) For DA evaluation only, use 1, 2, or 3 standard deviations of trait distribution to create user-defined "training samples". For molecular data: (1) Transform raw marker data to identify individual alleles; (2) Fill in missing data using Multiple Imputation (SAS Institute, ver. 9.1); (3) Identify subpopulations by model-based method (www.stats.ox.ac.uk/~pritch/home.html); (4) Perform AMOVA (Excoffier et al., 1992) of marker profiles to test differences between pre-defined groups using Arlequin software (Achneider et al., 2002); (5) Perform discriminant analysis (proc

STEPDISC, SAS Institute, ver. 9.1) to identify marker(s) that best differentiate training samples; (7) Use newly released SAS GLMSELECT to perform stepwise forward regression with a 0.05 significance level and the smallest CVPRESS as criteria to select up to 20 alleles (proc GLMSELECT, SAS institute, ver. 9.1); (8) Perform LASSO and LAR with the smallest CVPRESS as a criterion to select up to 20 alleles (proc GLMSELECT, SAS institute, ver. 9.1); (9) Perform two-way interaction selections with each GLMSELECT selected model, specifying that interactions can enter the model only if the corresponding main effects are already in the model (proc GLMSELECT, SAS institute, ver. 9.1); (10) Calculate R^2 values with proc REG (SAS Institute, ver. 9.1); (11) Evaluate goodness of fit for different models with proc MIXED (SAS Institute, ver. 9.1). SSR and RFLP markers were located on the Rice–Cornell SSR 2001-1 and /or Rice–Cornell RFLP 2001–2002 genetic maps (http://www.gramene.org).

3.3 Results and Discussion

DA, stepwise forward regression, and LASSO/LAR models identified the same and different alleles for all traits (Table 3.1). Those alleles identified by all three approaches should be considered as good candidates for additional analysis and fine mapping studies. DA-selected markers collectively explained 54% to 93% of the observed variation for 12 traits while individual alleles explained 0.01% to 63% of the variation in training samples. Compared with LASSO/LAR, stepwise forward regression markers explained a relatively high proportion of the observed variation for all the 12 traits (60%-94% vs. 48%-87%). However, LASSO/LAR tended to select markers with higher individual R^2 than stepwise forward regression. Similar results were obtained by LASSO and LAR most likely because LAR is a solution of LASSO.

Table 3.1. SSR/RFLP alleles identified by discriminant analysis (1 SD), stepwise forward regression, and LASSO/LAR regression with R^2 values in rice.

Trait	Discriminant Analysis			Stepwise Forward Regression			LASSO/LAR		
	Allele	Individual R^2	Total R^2	Allele	Individual R^2	Total R^2	Allele	Individual R^2	Total R^2
Tiller number	RZ400_15_2 ΔΨ [a, b, c]	.5638	.7175	RM212_116_4 ΔΨ	.4874	.9432	RM212_116_4 ΔΨ	.4874	.8716
	RM212_116_4 ΔΨ	.5330		RM48_215_5 Ψ	.4339		RM212_114_3 Ψ	.4785	
	RM14_187_7	.0024		RM202_158_3 Ψ	.3769		RM262_141_1	.4754	
	RM1_95_9	.0084		RZ400_15_2 ΔΨ	.3377		RM48_215_5 Ψ	.4339	
	RZ404_33_2	.0049		RM212_114_3 Ψ	.4785		RM26_114_3	.4235	
	RM259_162_11	.0054		RM228_150_18	.1446		RM202_158_3 Ψ	.3769	
	RM204_140_16	.0104		RM219_246_21	.0134		RM48_205_1	.4225	
	RM226_193_1	.0049		RG480_48_2 Ψ	.3958		RG480_48_2 Ψ	.3958	
	RM13_151_10	.0003		RZ537A_64_8	.0215		RZ400_15_2 ΔΨ	.3377	
	RM240_132_5	.0098		RM21_148_6 Ψ	.1673		RM226_195_2	.3180	
	RZ599_78_5	.0112		RM38_244_6	.0960		RM262_157_6	.3432	
	RM48_211_4	.0004		RM226_197_3	.0020		RM259_156_5	.3359	
	RZ141A_70_3	.0007		RM263_164_6	.0258		RM211_155_4	.2421	
	RM257_170_21	.0071		RM257_150_10	.0254		RM21_148_6 Ψ	.1673	
	RM253_140_15	.0009		RM226_262_16	.0058		RM224_157_12	.3210	
Plant height (cm)	CDO405_140_3 Δ	.0871	.7191	RM255_151_7 Ψ	.1063	.7252	RM255_151_7 Ψ	.1064	.5993
	CDO580_6_1 Δ	.0911		RM212_136_7 ΔΨ	.0392		RZ284_94_2	.0915	
	RZ424B_26_1 ΔΨ	.0660		RM21_132_2 ΔΨ	.0871		RM263_156_2 ΔΨ	.0903	
	RM212_136_7 ΔΨ	.0640		RM263_156_2 ΔΨ	.0903		RM21_132_2 ΔΨ	.0871	
	RM263_156_2 ΔΨ	.1102		RM259_174_20 Ψ	.0414		RM224_135_4	.0833	
	RM247_154_12	.0161		RM228_124_10	.0198		RZ251_200_4	.0849	
	RM21_132_2 ΔΨ	.1236		RM14_193_10	.0188		RM248_80_3 Ψ	.0584	
	RM235_96_2	.0152		RZ424B_26_1 ΔΨ	.0415		RM212_136_7 ΔΨ	.0392	
	RM1_89_6	.0137		RM248_80_3 Ψ	.0584		RM259_174_20 Ψ	.0414	
	RG716_154_4	.0264		RM207_123_9	.0001		RZ424B_26_1 ΔΨ	.0415	
	RM232_156_10	.0281		RM23_148_7	.0070		CDO405_140_3 Δ	.0458	
	RM22_187_3	.0033		CDO580_6_1 Δ	.0598		CD0395_75_3	.0478	
	RZ53_22_1 Δ	.0525		RZ143_63_1	.0108		RZ53_22_1 Δ	.0342	
	RM232_150_7	.0185		RM55_229_2	.0323		RM259_155_4	.0489	
	RM205_155_12	.0832		RM14_173_3	.0235		RM207_129_12	.0246	
Heading date (d)	RM263_156_2 ΔΨ	.1063	.5362	RM209_117_2 Ψ	.1125	.7283	RM209_117_2 Ψ	.1125	.4700
	RM248_92_8 Δ	.1004		RM18_157_3 Ψ	.1106		RM18_157_3 Ψ	.1106	
	RM255_149_6 Δ	.0309		RG716_96_3 ΔΨ	.0470		RM202_158_3	.0631	
	RZ740_60_5	.0055		RM10_169_6 Ψ	.0550		RM47_232_2	.0736	
	RG716_96_3 ΔΨ	.0949		RG64_37_2	.0005		RM263_156_2 ΔΨ	.0681	
	RM219_210_9	.0040		RM224_157_12	.0440		RG716_96_3 ΔΨ	.0470	
	RG146_31_1	.0024		RM263_156_2 ΔΨ	.0681		RM247_134_4	.0635	
	RM204_178_30	.0209		RM204_144_18	.0368		RM248_92_8 Δ	.0771	
	RM255_141_2	.0014		RM47_230_1	.0143		RM235_134_14	.0604	
	RM44_130_13	.0053		RM204_180_31	.0222		RM19_246_7	.0684	
	RM224_140_7	.0030		RM219_244_20	.0407		RM10_169_6 Ψ	.0550	
	RM55_231_3	.0000		RM10_177_11	.0000				
	RM22_193_6	.0133		RM207_141_17	.0009				
	RZ53_22_1	.0004		RM255_149_6 Δ	.0229				
	RZ537A_64_8	.0002		RM241_140_14	.0052				
1000-grain weight (g)	RM248_82_4 Δ	.2112	.6494	CDO365_160_4 Ψ	.1390	.6852	CDO365_160_4 Ψ	.1390	.5079
	RM205_161_15 ΔΨ	.0956		RM204_130_13 Ψ	.0465		RM248_82_4 Δ	.1378	
	RG901_144_4	.1249		RZ596_260_5 Ψ	.0894		RM228_116_6	.1228	
	RM223_147_3	.0834		RM44_108_6 Ψ	.0853		RZ596_260_5 Ψ	.0894	
	RM204_178_30 Δ	.0105		RM205_161_15 ΔΨ	.0719		RZ424B_210_7	.0840	
	RM205_153_11 Δ	.0105		RM205_153_11 Δ	.0127		RM44_108_6 Ψ	.0853	
	RM224_138_6	.0015		RM228_138_14	.0373		RM232_160_12	.0945	
	RM259_159_8	.0075		RM204_178_30 Δ	.0051		CDO38_41_3	.0989	
	RM38_266_17	.0038		RM55_219_1	.0009		RM51_132_2	.1130	
	RM226_221_10	.0066		RZ892_196_4	.0435		RM205_161_15 ΔΨ	.0719	
	RM205_127_6	.0064		BCD808A_74_4	.0026		RM248_86_6 Ψ	.0814	
	RM226_273_18	.0059		RZ206B_66_1	.0001		RM1_91_7	.0859	
	RM1_85_4	.0025		RM257_191_34	.0446		RM7_175_7	.0955	
	CDO20_94_1	.0177		RM248_86_6 Ψ	.0814		RM204_130_13 Ψ	.0465	
	RM202_156_1	.0046		RG20_140_1	.0034		RM30_78_1	.0798	

(**Table 3.1** continued)

Grain length (L) (mm)	RZ574_215_2 ΔΨ	.5252	.8133	RZ574_215_2 ΔΨ	.3851	.8178	RZ574_215_2 ΔΨ	.3851	.7209
	RM16_184_6	.0148		RM11_127_3 Ψ	.2991		RM11_127_3 Ψ	.2991	
	RG757_150_2 Δ	.4563		RM10_173_9 Ψ	.0563		RG757_150_2 Δ	.3205	
	RM19_237_6 Δ	.0087		RM259_159_8 Ψ	.0400		RM14_183_6	.2546	
	RM20B_207_1	.1086		RM55_229_2 Ψ	.0520		RM235_96_2	.2003	
	RM1_93_8	.0039		RM228_120_8 ΔΨ	.0277		RG64_37_2	.2783	
	RM228_120_8 ΔΨ	.0251		RM204_156_22	.0894		RZ103B_97_2	.2271	
	RM1_117_16	.0020		RM259_175_21 Ψ	.0342		RM239_143_2 Ψ	.1504	
	RM205_127_6 Δ	.0026		RM19_237_6 Δ	.0046		RM10_173_9 Ψ	.0563	
	RM239_144_3	.0024		RM21_168_16	.0001		RM11_133_5	.2578	
	RM257_185_31	.0039		RM207_123_9	.0008		RM226_203_4	.0447	
	RM14_187_7	.0001		RM211_143_1	.0229		RM259_159_8 Ψ	.0400	
	RZ537A_26_1	.0135		RZ599_78_5 Δ	.0173		RM55_229_2 Ψ	.0520	
	RM219_212_10	.0043		RM205_127_6 Δ	.0002		RM228_120_8 ΔΨ	.0277	
	RZ599_78_5 Δ	.0192		RM239_143_2 Ψ	.1504		RM259_175_21 Ψ	.0342	
Grain width (W) (mm)	RM248_82_4 Δ	.5237	.8680	CDO365_160_4 Ψ	.3568	.8605	CDO365_160_4 Ψ	.3568	.7453
	RM202_159_4	.3633		RM248_86_6 Ψ	.3132		RM248_82_4 Δ	.3567	
	RM14_183_6 ΔΨ	.5222		RM14_183_6 ΔΨ	.3241		RZ400_32_3 Δ	.3279	
	RM51_132_2 ΔΨ	.4366		RM223_149_4 Ψ	.2017		RM248_86_6	.3132	
	RM263_184_17 Δ	.0147		RM51_132_2 ΔΨ	.3171		RM14_183_6 ΔΨ	.3241	
	RZ625_180_4	.0671		RZ424B_210_7	.0464		RM209_127_4	.3028	
	RM232_164_14	.0124		RM263_184_17 Δ	.0132		RM51_132_2 ΔΨ	.3171	
	RM247_162_16	.0017		RM232_158_11	.1350		RM21_160_12	.2981	
	RM247_172_20	.0199		RM250_160_6	.0296		RM232_160_12	.2759	
	RZ599_40_3	.0283		RM1_95_9	.0115		RM223_149_4 Ψ	.2017	
	RZ400_32_3 Δ	.4402		RM241_128_6	.0391		RM10_175_10 Δ	.2275	
	RM207_117_6	.0708		RM263_162_5	.0508		RM235_100_4	.2094	
	RM10_175_10 Δ	.2842		RM1_89_6 Ψ	.2140		RM1_89_6 Ψ	.2140	
	RZ783_40_2	.2342		CDO38_39_2	.0295		CDO87_35_3	.1746	
Grain L/W ratio	RZ574_215_2 ΔΨ	.6281	.9308	RM209_135_8 Ψ	.3918	.8661	RM209_135_8 Ψ	.3918	.7511
	RM258_150_7 Δ	.0373		RM14_183_6 Ψ	.3407		RM21_160_12	.3850	
	RM21_162_13	.0163		RM232_160_12 Ψ	.2765		RM14_183_6 Ψ	.3407	
	RM21_154_9	.0443		RM10_173_9	.0466		RM11_127_3	.3300	
	RM202_184_12	.0049		RM13_147_8	.0256		RZ574_215_2 ΔΨ	.3605	
	RM209_161_16	.0080		RZ574_215_2 ΔΨ	.3605		RM209_127_4	.3359	
	RZ405_158_8	.0066		RM258_150_7 Δ	.0187		RM248_82_4	.2935	
	RM240_132_5	.0280		RM30_86_3	.0177		RZ103B_97_2	.3295	
	RM10_166_4	.0035		RM263_184_17	.0113		RM10_175_10	.3078	
	RM226_269_17	.0095		RM1_95_9	.0280		RM232_160_12 Ψ	.2765	
	RM262_143_2	.0030		RM51_132_2 Ψ	.2641		RM51_132_2 Ψ	.2641	
	RM13_151_10	.0001		RM223_149_4 Ψ	.1786		RM223_149_4 Ψ	.1786	
	RZ599_78_5	.0153		RM223_157_8	.0134		RM235_96_2	.2386	
	RM226_219_9	.0034		RM27_156_3	.0132		RM55_229_2	.0955	
Grain thickness (mm)	ND[d]	ND	ND	RM14_183_6 Ψ	.3896	.8201	RM14_183_6 Ψ	.3896	.7356
				RM223_149_4 Ψ	.2428		RZ400_32_3	.3682	
				RM232_160_12 Ψ	.1907		RZ87_19_1	.3545	
				RM235_100_4 Ψ	.2277		CDO365_160_4	.3228	
				RM209_127_4 Ψ	.3009		RM248_82_4	.3169	
				CDO20_94_1 Ψ	.1066		RM209_127_4 Ψ	.3009	
				RM47_230_1	.0219		RM223_149_4 Ψ	.2428	
				RZ599_38_2	.0091		RM248_86_6	.2180	
				RM263_179_13	.0008		RG322A_16_1	.3125	
				RZ87_22_2	.3313		RM51_132_2	.2564	
				RM250_160_6	.0201		RM235_100_4 Ψ	.2277	
				RZ323_49_2	.0017		RM232_160_12 Ψ	.1907	
				RM202_156_1	.0024		CDO20_94_1 Ψ	.1066	
				RZ206B_66_1	.0184		RM253_132_10	.1678	
				RM219_206_7	.0880		RM228_116_6	.0786	
Flag leaf length (cm)	CDO456_36_6 Δ	.1621	.5624	RM51_138_5 Ψ	.0885	.6029	RM51_138_5 Ψ	.0885	.4780
	RZ424B_26_1	.0119		RM230_237_1 Ψ	.0806		RZ424A_11_1 Δ	.0840	
	RM235_136_15	.0147		RM204_144_18 ΔΨ	.0519		RM230_237_1 Ψ	.0808	
	RM19_222_4	.0090		RM211_163_5 Ψ	.0409		RM10_169_6 Ψ	.0806	
	RM263_160_4	.0434		RM10_169_6 Ψ	.0701		CDO456_36_6 Δ	.0744	
	RM230_259_6	.0068		RM232_158_11 Ψ	.0586		RM204_144_18 ΔΨ	.0701	
	RM257_173_24	.0094		BCD386_91_2 Ψ	.0334		RM209_117_2	.0657	
	RM204_144_18 ΔΨ	.0647		RM204_114_6 Ψ	.0300		RM232_158_11 Ψ	.0646	
	CDO202_250_4	.0023		RM258_150_7	.0175		RM263_156_2	.0586	
	RM223_161_10	.0084		RM204_174_28	.0052		RM211_163_5 Ψ	.0552	
	RZ424A_11_1 Δ	.1302		RM215_156_7 Ψ	.0503		BCD386_91_2 Ψ	.0509	
	RM228_152_19	.0156		RM38_266_17 Ψ	.0253		RM215_156_7 Ψ	.0382	
	RG716_154_4	.0108		RZ740_58_4	.0044		RM38_266_17 Ψ	.0310	
	RM224_132_2	.0105		RZ649_21_1	.0098		RM204_114_6 Ψ	.0300	

(**Table 3.1** continued)

Flag leaf width (cm)	RM44_130_13 Δ	.1486	.6128	RM11_127_3 Ψ	.1170	.7545	RM11_127_3 Ψ	.1170	.6583
	CDO78_63_2	.0298		RM22_189_4 Ψ	.0662		RM44_130_13 Δ	.1062	
	RM44_120_12	.0155		RM14_169_1 Ψ	.0649		RM232_162_13 Ψ	.0969	
	RM215_148_3 ΔΨ	.0721		RM232_162_13 Ψ	.0969		RM14_169_1 Ψ	.0649	
	RZ599_38_2 Δ	.0919		RM10_171_8 Ψ	.0593		RM22_189_4 Ψ	.0662	
	RM44_112_8	.0058		RM207_119_7 Ψ	.0650		RM11_123_1	.0730	
	RZ53_22_1 Δ	.0001		RZ390_94_3	.0022		RM10_171_8 Ψ	.0593	
	RM233B_142_5 ΔΨ	.0551		RM215_148_3 ΔΨ	.0580		RZ599_38_2 Δ	.0673	
	RM239_144_3	.0096		CDO718_39_1 Ψ	.0563		RZ421_19_1	.0648	
	RM38_250_9	.0092		RM230_257_5	.0266		RM207_119_7 Ψ	.0650	
	RM255_143_3	.0133		RM233B_142_5 ΔΨ	.0439		RM215_148_3 ΔΨ	.0580	
	BCD98_44_3	.0151		RM14_171_2 Δ	.0122		RM259_155_4 Ψ	.0556	
	RM14_189_8	.0090		RZ53_22_1 Δ	.0004		CDO718_39_1 Ψ	.0563	
	RM14_171_2 Δ	.0139		RM205_153_11 Ψ	.0437		RM233B_142_5 ΔΨ	.0439	
	RM10_179_12	.0029		RM259_155_4 Ψ	.0556		RM205_153_11 Ψ	.0437	
Stem diameter (mm)	RM255_151_7 Δ	.1587	.5701	RM44_130_13 Ψ	.1126	.6973	RM44_130_13 Ψ	.1126	.5107
	RZ424A_11_1 ΔΨ	.1548		RZ424A_11_1 ΔΨ	.1059		RM255_151_7 Δ	.1091	
	RM222_213_9	.0270		RZ405_158_8 Ψ	.0587		RZ424A_11_1 ΔΨ	.1059	
	RZ284_67_1	.0806		RM224_158_13 Ψ	.0851		RZ400_32_3	.1080	
	RM19_216_2	.0424		RZ886_48_3 Ψ	.0736		RM232_158_11	.0969	
	RM235_136_15	.0360		RZ537A_26_1	.0333		RZ395_44_1	.0991	
	RM222_209_7	.0342		RM262_157_6	.0264		RM224_158_13 Ψ	.0851	
	CDO405_180_5	.0110		RM259_155_4 Ψ	.0425		RM263_156_2	.0863	
	RM21_164_14	.0027		RM20B_210_2	.0128		RZ87_19_1 Ψ	.1044	
	RZ405_160_9	.0572		RM26_114_3	.0380		RM232_156_10	.0695	
	CDO98_50_1	.0183		RZ143_91_2 Δ	.0019		RZ405_158_8 Ψ	.0587	
	RM27_158_4	.0079		RZ87_19_1 Ψ	.1044		RZ886_48_3 Ψ	.0736	
	RZ143_91_2 Δ	.0040		RM232_150_7 Ψ	.0532		RM232_150_7 Ψ	.0532	
	RZ103A_54_3	.0199		RM21_132_2	.0497		RM259_155_4 Ψ	.0425	
Panicle length (cm)	RM7_175_7 ΔΨ	.2740	.6174	CDO405_170_4 ΔΨ	.1876	.6513	CDO405_170_4 ΔΨ	.1876	.5267
	RM232_144_4	.0341		RZ2_47_2	.0768		RM7_175_7 ΔΨ	.1876	
	RZ424B_26_1	.0033		RM14_173_3 Ψ	.0781		RM14_183_6	.1658	
	RM224_134_3	.0195		RM11_129_4 Ψ	.0612		RM235_100_4	.1407	
	RM20A_276_10 Δ	.0975		RM7_175_7 ΔΨ	.1867		RM228_114_5	.1386	
	RM263_160_4 ΔΨ	.0412		RM259_174_20 Ψ	.0176		RM209_127_4	.1287	
	RM212_112_2 Δ	.0376		RM22_195_7	.0012		RZ2_45_1	.0768	
	RZ141B_240_2	.0052		RZ387_12_1	.0328		RM11_129_4 Ψ	.0612	
	CDO405_170_4 ΔΨ	.2600		RM226_219_9	.0119		RM14_173_3 Ψ	.0781	
	RM38_266_17	.0142		RM263_160_4 ΔΨ	.0309		RM20A_276_10 Δ	.0747	
	RG716_86_2	.0196		RZ424B_190_6	.0443		RM263_160_4 ΔΨ	.0309	
	RM235_96_2	.0834		RM212_112_2 Δ	.0338		RM259_174_20 Ψ	.0176	

[a] First component of designation is SSR/RFLP marker, second is allele size in bp (SSR) or 100 bp (RFLP), and third is allele number at locus. Allele order in table corresponds to its relative contribution to the model.

[b] Alleles with Δ sybmol were identified by both DA and stepwise forward regression or LASSO/LAR regression.

[c] Alleles with Ψ symbol were identified by both stepwise forward regression and LASSO/LAR regression.

[d] No alleles identified most likely due to lack of phenotypic variation observed among lines.

Analysis with population structure increased R^2 values for all traits except for tiller number with stepwise forward regression and flag leaf width with LASSO/LAR selection (Table 3.2). The epistatic model containing interaction terms and fewer main effects increased R^2 values for all traits except for 1000-grain weight and panicle length with LASSO/LAR selection, grain thickness and flag leaf length with LASSO/LAR selection, and stem diameter with stepwise

forward regression. Stepwise forward regression and LASSO/LAR models identified the same and different interaction terms for all traits. Some interaction terms (RM48_215_5*RM202_158_3, RM212_114_3*RG480_48_2 and RM212_114_3*RM226_197_3 of tiller number, RM232_160_12*RM10_175_10 of grain width, RM209_135_8* RM10_175_10 and RM10_175_10* RM11_133_5 of grain length-width ratio, RM209_135_8*CDO20_94_1 of grain thickness) explained a very high proportion of the observed variation (>40%) (Table 3.2).

Models with interaction terms showed a significant improvement (at level $p < 0.001$) in goodness of fit compared with the models without interaction terms for all traits except for plant height, grain length, grain thickness and flag leaf length (Table 3.3). When significant *p* levels were adjusted to < 0.05, those four cases were also significant. Potential epistatic loci with high R^2 values were identified that merit further investigation.

Markers selected by DA/stepwise forward regression/LASSO/LAR mapped within known QTLs for agronomic traits. For example, DA-selected allele CDO118_69_1 mapped within the 4.25 cM QTL *gw1.1* and 3.15 cM QTL *gw1.2* on chromosome 1 for 1000-grain weight (Moncada et al. 2001) (Fig. 3.1). It was also found 1.9 cM from the 18 cM QTL *QKw1* (Li et al. 1997). DA-selected allele RZ329_43_3 DA was detected 7.4 cM from the 11 cM QTL *QKw3a* (Li et al. 1997), 3.1 cM, 4.05 cM, and 5.65 cM from the unnamed loci affecting 1000 grain weight, respectively (Li et al. 1997), and 7.95 cM from the 48.9 cM QTL *gw3* (Xiao et al. 1996) on chromosome 3. SFR/LASSO/LAR-selected allele RM232_160_12 was observed within the QTL *gw3*. Allele RM241_138_13 selected by DA and RM241_142_15 selected by SFR/LASSO/LAR were found within the 75 cM unnamed QTL detected by Brondani et al. (2002) on chromosome 4. DA-selected Allele RM44_92_2 and SFR/LASSO/LAR selected allele RM44_108_6 mapped within the 32.9 QTL *gw-8* on chromosome 8 (Lu et al. 1997). Moreover,

Table 3.2. Epistatic loci and R^2 values determined for twelve rice traits derived from modified stepwise forward and LASSO/LAR regression methods with or without consideration of population structure.

Trait	Method	Assuming no structure					Assuming structure (Sub population 1)[a]				
		Additive		Epistatic			Additive		Epistatic		
		Individual R^2 range	Total R^2	Interaction term	Individual R^2	Total R^2	Individual R^2 range	Total R^2	Interaction term	Individual R^2	Total R^2
Tiller number	Stepwise Forward	.0001~.4874	.9432	RM48_215_5*RM202_158_3 ΔΨ[bc] RM212_116_4*RZ400_15_2	.4134 .3245	.9439	.0106~.6130	.9244	RM212_114_3*RM253_136_13 Δ RM226_197_3*RM253_136_13 Δ	.2359 .3082	.9345
	LASSO/LAR	.1446~.4874	.8716	RM48_215_5*RM202_158_3 ΔΨ RM26_114_3*RM211_155_4 RM212_114_3*RG480_48_2 Ψ	.4134 .2905 .4071	.9037	.0229~.6130	.9182	RM212_114_3*RM226_197_3 Ψ RM212_114_3*RM253_136_13 Δ RM226_197_3*RM253_136_13 Δ	.5056 .2359 .3082	.9200
Plant height (cm)	Stepwise Forward	.0001~.1063	.7252	RM255_151_7*RM21_132_2 Δ RM263_156_2*CDO580_6_1 RM14_193_10*RM23_148_7	.0136 .1116 .0489	.7719	.0011~.1561	.8029	RZ103A_54_3* RM248_80_3 Δ RM263_156_2* RM22_195_7 RM263_156_2* RM202_182_11	.0480 .0917 .0224	.8451
	LASSO/LAR	.0246~.1064	.5993	RM255_151_7*RM21_132_2 Δ RM263_156_2*CD0395_75_3 RM21_132_2*RM259_155_4	.0136 .0144 .0723	.6645	.0585~.1561	.6844	RZ103A_54_3* RM248_80_3 Δ RM263_156_2* RM255_151_7 RM255_151_7* RM20B_210_2	.0480 .2126 .0566	.7065
Heading date (d)	Stepwise Forward	.0001~.1125	.7283	RG64_37_2*RM204_180_31 RG64_37_2*RM241_140_14	.0407 .0407	.7531	.0001~.1215	.7714	RM48_215_5* CDO87_19_1	.0055	.7879
	LASSO/LAR	.0550~.1125	.4700	RG716_96_3*RM19_246_7	.0407	.4737	.0302~.1212	.5716	RM263_156_2* RM204_144_18 RM212_114_3* CDO38_41_3	.0055 .0196	.5884
1000-grain weight (g)	Stepwise Forward	.0001~.1390	.6852	BCD808A_74_4*RG20_140_1 RZ596_260_5* RM55_219_1 RM205_161_15*RZ892_196_4	.0126 .0417 .0615	.7382	.0001~.2268	.8192	RM44_108_6* RM207_125_10 RM241_142_15*RM10_173_9 RM44_108_6* RM21_156_10	.0077 .0104 .0093	.8319
	LASSO/LAR	.0414~.1390	.5079	RZ596_260_5* RZ424B_210_7	.0073	.4894	.0286~.2268	.6450	RM228_116_6* CD020_164_2 RM232_160_12*CD020_164_2	.0546 .0181	.6542
Grain length (L) (mm)	Stepwise Forward	.0001~.3851	.8178	RM11_127_3*RM55_229_2 RM11_127_3*RM207_123_9	.0001 .0304	.8215	.0001~.5021	.8915	RM10_173_9*RZ251_180_3	.0025	.8953
	LASSO/LAR	.0277~.3851	.7209	RZ574_215_2*RM55_229_2 RM235_96_2*RM55_229_2 RM235_96_2*RM259_175_21	.0164 .0496 .0420	.7359	.0505~.5021	.8179	RM14_183_6*RG64_37_2 RZ574_215_2*RM230_255_4 RG64_37_2*RM230_255_4	.3617 .2627 .2623	.8420
Grain width (W) (mm)	Stepwise Forward	.0018~.3568	.8605	RM14_183_6*CDO38_39_2	.0384	.8654	.0007~.5890	.9182	RM257_171_22*RZ886_29_1 RM223_149_4*RM228_116_6	.0385 .0369	.9281
	LASSO/LAR	.0303~.3568	.7453	RM14_183_6*RM51_132_2 RM235_100_4*CDO87_35_3	.2998 .2454	.7592	.0339~.5890	.8503	RM10_175_10*RM223_149_4 RM232_160_12*RM10_175_10 Ψ	.0258 .4073	.8535

(**Table 3.2** continued)

Grain L/W ratio	Stepwise Forward	.0010~.3918	.8661	RM232_160_12*RM51_132_2 RM14_183_6* RM263_184_17 RM223_157_8* RM27_156_3	.0172 .0054 .0067	.8780	.0003~.6089	.9360	RM232_160_12*RM51_132_2 Δ RM232_160_12*RG322A_16_1 Ψ RZ574_215_2* RM10_173_9	.0344 .4176 .1488	.9505
	LASSO/LAR	.0955~.3918	.7511	RM11_127_3* RM248_82_4 RM209_135_8* RM10_175_10 Ψ RM209_135_8* RZ103B_97_2	.2767 .4247 .2947	.7857	.0338~.6089	.8720	RZ574_215_2* RM223_149_4 RM232_160_12*RM51_132_2 Δ RM10_175_10* RM11_133_5 Ψ	.2479 .0344 .5473	.8765
Grain thickness (mm)	Stepwise Forward	.0001~.3896	.8201	RM14_183_6*RM235_100_4 Δ RZ206B_66_1*RM219_206_7	.2997 .0975	.8486	.0001~.5051	.8874	RM223_149_4*RM224_134_3 RM209_135_8*CDO20_94_1 Ψ	.0092 .5169	.8983
	LASSO/LAR	.0442~.3896	.7356	RM14_183_6*RM235_100_4 Δ RM223_149_4*RM253_132_10	.2997 .1614	.7524	.2814~.5051	.7604	RM209_135_8*RM44_92_2	.3277	.7555
Flag leaf length (cm)	Stepwise Forward	.0003~.0885	.6029	RZ740_58_4*RZ649_21_1 RM204_144_18*RZ740_58_4	.0243 .0632	.6298	.0002~.0886	.6946	RM204_144_18*CDO545A_102_2 RZ400_32_3*RM259_173_19	.0972 .0275	.7399
	LASSO/LAR	.0253~.0885	.4780	ND[d]	ND	ND	.0276~.0886	.5361	RM55_229_2*RM263_156_2	.0164	.5184
Flag leaf width (cm)	Stepwise Forward	.0004~.1170	.7545	RM11_127_3*CDO718_39_1 Δ RM22_189_4*RM232_162_13 RM207_119_7*RM230_257_5	.0657 .0629 .0245	.7814	.0001~.1965	.7761	RM44_130_13*RM207_119_7 Δ RM22_189_4*CDO718_41_2 RM232_162_13*RZ599_38_2	.1359 .0097 .0881	.8019
	LASSO/LAR	.0218~.1170	.6583	RM232_162_13*RM233B_142_5 RM11_127_3*CDO718_39_1 Δ RZ599_38_2*RZ421_19_1	.0629 .0657 .0715	.6796	.0755~.1965	.6364	RM44_130_13*RM207_119_7 Δ RM44_130_13*RZ421_19_1 RM22_189_4*CDO718_39_1	.1359 .0392 .1120	.6749
Stem diameter (mm)	Stepwise Forward	.0019~.1126	.6973	RM224_158_13*RM20B_210_2 RM259_155_4*RM21_132_2 RM224_158_13*RZ886_48_3 Δ	.0990 .0596 .1191	.7320	.0002~.1658	.7910	RM224_158_13*RZ886_29_1 Δ RZ143_63_1*RM241_142_ 15 RZ886_29_1*RM11_127_3	.0218 .0498 .1110	.7753
	LASSO/LAR	.0425~.1126	.5107	RZ424A_11_1*RM259_155_4 RM224_158_13*RZ886_48_3 Δ	.0001 .1191	.5209	.0589~.1658	.6618	RM224_158_13*RM20B_210_2 RM224_158_13*RZ886_29_1 Δ	.2291 .0218	.6731
Panicle length (cm)	Stepwise Forward	.0012~.1876	.6513	RM22_195_7*RZ424B_190_6 RM7_175_7*RZ424B_190_6	.0008 .0239	.6653	.0001~.2171	.7414	CDO405_170_4*RM20A_305_20 CDO98_65_3*RZ206B_66_1	.0008 .0428	.7689
	LASSO/LAR	.0185~.1876	.5267	CDO405_170_4*RM7_175_7	.1453	.5258	.0364~.2171	.5506	RM14_173_3*RM223_147_3 RM14_183_6*RM235_96_2	.2220 .0070	.5550

[a] Only sub-population 1 was evaluated because the remaining two subpopulations contained insufficient size (n = 16, 43) for analysis

[b] Interaction terms with Δ sybmol were identified by both stepwise forward regression and LASSO/LAR regression.

[c] Interaction terms with Ψ symbol had large effects with R^2 >.4000.

[d] No Interaction terms were selected.

Table 3.3. Goodness of fit of two different models in explaining phenotypic variation for twelve rice traits.

	Tiller number		Plant height		Heading date		1000-grain weight	
Model [a]	-2 log likelihood	BIC[b]	-2 log likelihood	BIC	-2 log likelihood	BIC	-2 log likelihood	BIC
Without interaction	530.6	535.6	-244.8	-239.8	837.9	842.9	555.4	560.4
With interaction	497.3*[c]	502.3	-249.8 (NS)	-244.9	807.6*	812.5	534.1*	539.0

(**Table 3.3** continued)

	Grain length (L)		Grain width (W)		Grain L/W ratio		Grain thickness	
Model	-2 log likelihood	BIC	-2 log likelihood	BIC	-2 log likelihood	BIC	-2 log likelihood	BIC
Without interaction	172.2	177.1	-107.6	-102.7	-1.4	3.5	-372.3	-367.3
With interaction	168.1 (NS)	173.0	-124.4*	-119.5	-31.5*	-26.6	-381.1 (NS)	-376.1

(**Table 3.3** continued)

	Flag leaf length		Flag leaf width		Stem diameter		Panicle length	
Model	-2 log likelihood	BIC	-2 log likelihood	BIC	-2 log likelihood	BIC	-2 log likelihood	BIC
Without interaction	752.8	757.8	-161.6	-156.7	226.2	231.1	569.9	574.9
With interaction	719.8*	724.7	-171.5 (NS)	-166.5	199.8*	204.7	542.0*	546.9

[a] Epistatic models with or without interaction terms in sub-population 1 were evaluated.

[b] BIC, Bayesian Information Criterion (smaller is better).

[c] Model comparison based on Chi-square test indicates whether the model with interaction terms significantly improved the model fit at $p < 0.001$; *, significant; NS, not significant.

the SFR/LASSO/LAR-selected alleles, RM232_160_12, RM241_142_15, and RM44_108_6 were one of the interaction terms for the trait of 1000-grain weight. For grain length / width ratio, DA-selected allele RM204_142_17 mapped within 31.4 cM unnamed QTL detected by Tan et al. (2000) on chromosome 6 (Fig. 3.1).

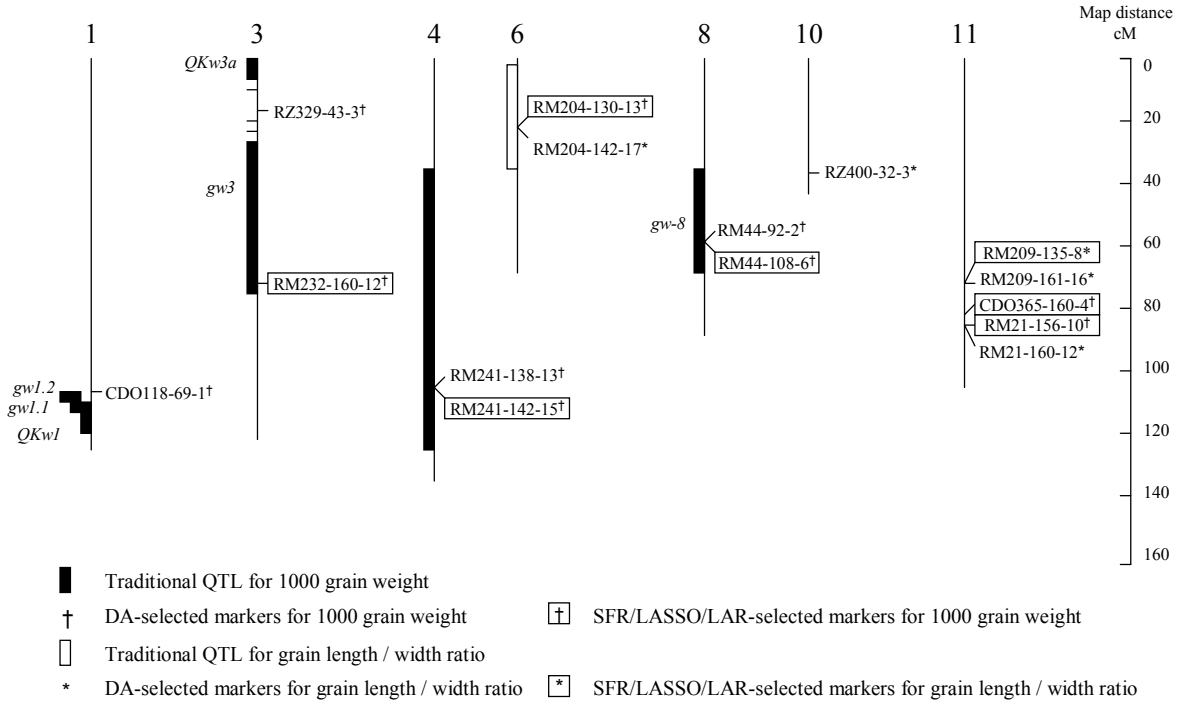

Fig. 3.1. Chromosomal locations of markers identified by discriminant analysis (DA), stepwise forward regression (SFR), least angle regression (LAR)/least absolute shrinkage and selection operator (LASSO), and quantitative trait locue (QTL) mapping for 1000-grain weight and grain length-width ratio in rice.

In addition to the DA/SFR/LASSO/LAR alleles that pointed to the same or nearby regions as previously reported QTLs, Fig. 3.1 shows several DA/SFR/LASSO/LAR-selected markers not found by traditional methods. For example, the following alleles selected for 1000-grain weight were not found associated with previously reported QTLs: SFR/LASSO/LAR alleles RM204_130_13 on chromosome 6 and RM21_156_10 and CDO365_160_4 on chromosome 11. Similarly, DA/SFR/LASSO/LAR alleles not associated with reported QTLs for grain length-width ratio include DA alleles RZ400_32_3 on chromosome 10, RM209_161_16 and RM21_160_12 on chromosome 11 and SFR/LASSO/LAR allele RM209_135_8 on

chromosome 11. These markers identified by DA/SFR/LASSO/LAR may therefore new loci associated with 1000-grain weight and grain length-width ratio. Similar results were obtained for the remaining ten agronomic traits evaluated in this study (data not shown).

3.4 Conclusions

The analysis with current statistical methods for association genetics including recently developed mixed models (Arbelbide et al. 2006; Yu et al. 2006), are conducted by one marker at a time independently. However, our current understanding of biochemical and physiological genetics, as well as the regulation of gene expression, strongly suggests the ubiquitous nature of interactions among gene products. Therefore, the complex trait such as yield and grain quality should be controlled by many genes and influenced by epistatic interactions of different loci and the approaches used in this study with the ability to simultaneously identify numerous loci and digenic epistatic interactions were reasonable and exhibited advantages compared to other approaches.

Results from this study indicated that the epistatic model based on stepwise forward regression successfully identified several interacting loci that explained a relatively high proportion of the observed variation for all the twelve agronomically important traits. Moreover, the loci identified by the epistatic model mapped within previously known QTL regions that underscores the genetic basis of the selected markers. Consideration of epistatic terms in the stepwise forward regression, along with population structure, and missing data (multiple imputation), created a more robust model compared to the one without epistasis as judged by log-likelihood and BIC comparisons. Results from this study suggest that association genetics is a rapid and powerful method to identify epistatic factors that impact agronomic traits of inbred

rice. The selected loci can be used in future studies to further dissect the contribution of epistasis to genetic variation and varietal improvement in rice.

3.5 References

Arbelbide M., Yu J., and Bernardo R. 2006. Power of Miexed-model QTL mapping from phenotypic, pedigree and marker data in self-pollinated crops. Theor Appl Genet 112:876-884

Brondani C., Rangel N., Brondani V., and Ferreira E. 2002. QTL mapping and introgression of yield-related traits from *Oryza glumaepatula* to cultivated rice (*Oryza sativa*) using microsatellite markers. Theor. Appl. Genet. 104:1192-1203

Cui K.H., Peng S.B., Xing Y.Z., Yu S.B. and Xu C.G. 2002. Genetic analysis of the panicle traits related to yield sink size of rice. Acta Genetica Sinica 29:144-152

Efron B., Hastie T., Johnstone I., and Tibshirani R. 2004. Least Angle Regression (with discussion). Annals of Statistics 32:407-499.

Fasoulas A.C. and Allard R.W. 1962. Nonallelic gene interactions in the inheritance of quantitative characters in barley. Genetics 47:899–907

Kraakman A.T.W., Niks R.E., Van den Berg P.M.M.M., Stam P., and Van Eeuwijk F.A. 2004. Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics 168:435-446

Kraja A.T. and Dudley J.W. 2000. QTL analysis of two maize inbred line crosses. Maydica 45:1-12

Lark K.G., Chase K., Adelf F., Mansur L.M. and Orf J.H. 1995. Interactions between quantitative trait loci in soybean in which trait variation at one locus is conditional upon a specific allele at another. Proc. Natl. Acad. Sci. USA 92:4656–4660

Li Z, Pinson S.R.M., Park W.D., Paterson A.H., and Stansel J.W. 1997. Epistasis of three grain yield components in rice (*Oryza sativa* L.). Genetics 145:453-465

Lu C.F., Shen L.H., Tan Z.B., Xu Y.B., He P., Chen Y., and Zhu L.H. 1997. Comparative mapping of QTLs for agronomic traits of rice across environments by using a doubled-haploid population. Theor. Appl. Genet. 94:145-150

Moncada P., Martinez C.P., Borrero J., Chatel M., Gauch H., Guimaraes E., Tohme J., and McCouch S.R. 2001. Quantitative trait loci for yield and yield components in an *Oryza sativa* x *Oryza rufipogon* BC_2F_2 population evaluated in an upland environment. Theor. Appl. Genet. 102:41-52

Okogbenin E. and Fregene M. 2002. Genetic analysis and QTL mapping of early root bulking in an F_1 population of non-inbred parents in cassava (*Manihot esculenta* Crantz). Theor. Appl. Genet. 106:58–66

Semagn, K., Bjornstad, A., Stedje, B., and Bekele, E. 2000. Comparison of multivariate methods for the analysis of genetic resource and adaptation in *Phytolacca dodecandra* using RAPD. Theor. Appl. Genet. 101:1145-1154

Storey J.D., Akey J.M., and Kruglyak L. 2005. Multiple locus linkage analysis of genomewide expression in yeast. PLoS Biology 3(e267):0001-0011

Tan Y.F., Xing Y.Z., Li J.X., Yu S.B., Xu C.G., and Zhang Q. 2000. Genetic bases of appearance quality of rice grains in Shanyou 63, an elite rice hybrid. Theor. Appl. Genet. 101:823-829

Tavoletti S., Pesaresi P., Barcaccia G., and Albertini E. 2000. Mapping the *jp* (jumbo pollen) gene and QTLs involved in multinucleate microspore formation in diploid alfalfa. Theor. Appl. Genet. 101:372-378

Tibshirani R. 1996. Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society Series B 58:267-288.

Virk P.S., Ford-Lloyd B.V., Jackson M.T., Pooni H.S., Clemeno T.P., and Newbury H.J. 1996. Predicting quantitative variation within rice germplasm using molecular markers. Heredity 76:296-304

Xiao J., Li J., Yuan L., and Tanksley S.R. 1996. Identification of QTLs affecting traits of agronomic importance in a recombinant inbred population derived from a subspecific rice cross. Theor. Appl. Genet. 92:230-244

Xing Y.Z., Tan Y.F., Hua J.P., Sun X.L., Xu C.G., and Zhang Q. 2002. Characterization of the main effects, epistatic effects and their environmental interactions of QTLs on the genetic basis of yield traits in rice. Theor. Appl. Genet. 105:248–257

Yu J., Pressoir G., Briggs W.H., Bi I.V., Yamasaki M., Doebley J.F., McMullen M.D., Gaut B.S., Holland J.B., Kresovich S., and Buckler E.S. 2006. A Unified Mixed-Model Method for Association Mapping Accounting for Multiple Levels of Relatedness. Nature Genetics 38:203-208

Zhang N., Xu Y., Akash M., McCouch S., and Oard J.H. 2005. Identification of candidate markers associated with agronomic traits in rice using Discriminant Analysis. Theor. Appl. Genet. 110:721-729

Zhuang J.Y. Fan Y.Y., Rao Z.M., Wu J.L., Xia Y.W., and Zheng K.L. 2002. Analysis on additive effects and additive-by-additive epistatic effects of QTLs for yield traits in a recombinant inbred line population of rice. Theor. Appl. Genet. 105:1137–1145

CHAPTER 4: CLONING AND PHYLOGENETIC ANALYSIS OF GENES ENCODING ACETOLACTATE SYNTHASE FROM *COREOPSIS TINCTRIA* NUTT. FOR RESISTANCE TO SULFONYLUREA AND IMIDAZOLINONE HERBICIDES

4.1 Introduction

Weeds are a constant limitation to optimal commercial crop production and can cause substantial yield losses in all growing areas of the world. The use of herbicides is considered an effective, easy and comparatively inexpensive approach to control noxious weeds. Acetolactate synthase (ALS) (acetohydroxyacid synthase, AHAS, E.C. 4.1.3.18), which catalyses the first common step in the biosynthesis of the branched-chain amino acids in plants, is a target of five herbicide groups, *viz.* sulfonylurea, imidazolinone, triazolopyrimidine, pyrimidinyl oxybenzoate, and sulfonylanminocarbonyl-triazolinones (Mallory-Smith and Retzinger, 2003). These herbicides block the biosynthesis of essential amino acids valine, leucine and isoleucine (Ray 1984; Santel et al. 1999; Schloss 1990; Shaner et al. 1984; Sibony and Rubin 2003; Stidham and Shaner 1990; Subrananian and Gerwick 1989; Subrananian et al. 1990). It is believed that starvation of plants for these amino acids is the primary mechanism by which ALS-inhibiting herbicides cause plant death (Tranel and Wright 2002). However, other secondary effects of ALS inhibition, such as buildup of α-ketobutyrate, disruption of protein synthesis, and disruption of photosynthate transport, have been implicated in the mechanism of plant death (Shaner 1991; Tranel and Wright 2002). ALS-inhibiting herbicides have been used as selective agents in laboratory studies to isolate a range of resistant biotypes from otherwise susceptible populations (Falco and Dumas 1985; Haughn and Somerville 1986).

4.1.1 ALS Enzyme and Mechanisms of Resistance to Herbicides

Two parallel reactions are catalyzed by the ALS enzyme: synthesis of (S)-2-acetolactate from two molecules of pyruvate and synthesis of (S)-2-aceto-2-hydroxybutyrate from a molecule

each of pyruvate and 2-ketobutyrate (Guttieri et al. 1996; McCourt et al. 2006; Singh et al. 1988). (S)-2-acetolactate is a precursor of valine and leucine while (S)-2-aceto-2-hydroxybutyrate is a precursor of isoleucine. In eukaryotes, ALS is encoded in the nucleus and is located in plastids of plants (Bowen et al. 1997; Duggleby and Pang 2000) or in mitochondria of fungi (Duggleby and Pang 2000). An N-terminal transit peptide is presumed to direct the protein to the appropriate organelle, and it is usually assumed that this transit peptide is cleaved during or after translocation. The site of cleavage has not yet been established for any ALS protein (Duggleby and Pang 2000). After Mazur et al. (1987) isolated the first two plant ALS genes from *Arabidopsis thaliana* (L.) Heynh. and *Nicotiana tabacum* L., a number of plant ALS genes have subsequently been cloned and characterized. Most diploid plant species have a single ALS locus (*Arabidopsis thaliana* (L.) Heynh. and *Xanthium strumarium* L.), with corn (*Zea mays* L.) and sunflower (*Helianthus annuus* L.*)* being notable exceptions with two loci (Guttieri et al. 1996) and three loci (Kolkman et al. 2004), respectively. Tetraploid tobacco (*Nicotiana tabacum* L.) has two loci (Chaleff and Bascomb 1987); *Brassica* species possess five loci (Rutledge et al. 1991); and *Gossypium Hirsutum* L. contain six loci (Grula et al., 1995). *Coreopsis tinctoria* Nutt. is reported to be a diploid species (Strother 1983), but the exact number of ALS loci is unknown. The mature ALS protein is approximately 670 amino acids long (Tan et al. 2005) and is highly conserved across species (Guttieri et al. 1996).

Certain related classes of herbicides are known to inhibit ALS, such as the sulfonylurea, imidazolinone, triazolopyrimidine, pyrimidinyl oxybenzoate, and sulfonylanminocarbonyl-triazolinones, by attaching to a relic quinine-binding site (Hattori et al. 1995, Santel et al. 1999). The discovery of ALS-inhibiting herbicides was a significant accomplishment in the history of weed science. A major advantage of these compounds is their high efficacy, broad-spectrum

weed control, low use rates, and environmental safety (Mazur and Falco 1989). Various sulfonylurea derivatives as potent herbicides were discovered in the mid-1970's (Levitt 1978), with action of these compounds due to inhibition of ALS in both plants (Chaleff and Mauvais 1984) and bacteria (LaRossa and Schloss 1984). American Cyanamid (now BASF Corp.) also developed the unrelated imidazolinone herbicides that also inhibit ALS (Shaner et al. 1984). Chlorsulfuron (Glean®) was the first ALS inhibitor marketed in North America in 1982 (Guttieri et al. 1996). Sulfometuron methyl (Oust®) is another sulfonylurea herbicide widely used (Guttieri et al. 1996). Imidazolinone herbicides, such as imazethapyr (Pursuit®) and imazaquin (Scepter®) are broad-spectrum weed control herbicides (Guttieri et al. 1996). Imazapic (Plateau®) is a selective herbicide for both the pre and post-emergent control of grasses, broad-leaf weeds, and weed control in natural areas, particularly in conjunction with the establishment of native warm-season prairiegrasses and certain legumes. Imazapic is relatively non-toxic to terrestrial and aquatic mammals, birds, and amphibians. Imazapic has an average half-life of 120 days in soil, and is rapidly degraded by sunlight in aqueous solution (Tu et al. 2001). Imazapyr (Arsenal®) is a non-selective herbicide used for the control of a broad range of weeds including annual and perennial grasses and broadleaved herbs, and woody species. Imazapyr is not highly toxic to birds and mammals, but some formulations can cause severe, irreversible eye damage. Studies indicate imazapyr is excreted by mammalian systems rapidly with no bioaccumulation (Tu et al. 2001). Worldwide, there are more than 30 commercial ALS-inhibiting herbicides, indicative of their importance for weed management in a wide range of crops (Shaner 1999).

The lack of inhibition of ALS in resistant plant biotypes is predominantly due to an altered form of ALS that is insensitive to certain herbicides (Christopher et al. 1992; Devine et al. 1991; Manley et al. 1999; Saari et al. 1990, 1992, 1994; Thill et al. 1993). A second

mechanism of resistance is enhanced herbicide metabolism resulting in rapid detoxification of the herbicide (Christopher et al. 1992; Saari et al. 1994; Veldhuis et al. 2000). Examples for the second mechanism include primisulfuron-tolerant corn (*Zea mays* L.) (Guttieri et al. 1996), chlorsulfuron-tolerant soybean (*Glycine max* (L.) Merr.) (Guttieri et al. 1996), rigid ryegrass (*Lolium rigidum* Gaud.) (Christopher et al. 1991; Cotterman and Saari 1992; Holtum et al. 1991), and blackgrass (*Alopecurus myosuroides* Huds.) (Kemp et al. 1990; Moss and Cussans 1991).

The crystal structure of the ALS enzyme has revealed a catalytic subunit formed by the folding of two large subunits that contain the A, B, C (Hershey et al. 1999; Lee and Duggleby. 2002; Pang et al. 2002) and D (Woodworth et al. 1996a) domains. Modeling research suggests that the binding site for ALS/AHAS herbicides resides near the junction between the two large subunits (Bekkaoui et al. 1993; Duggleby et al. 2003; Ott et al. 1996; Pang et al. 2002). Mutations for ALS resistance in the "pocket" of the folded ALS enzyme near the interface of the subunits were reported to act as a binding site for ALS/AHAS herbicides. Different classes of ALS/AHAS herbicides have been proposed to attach to unique, but overlapping regions of the binding site (McCourt et al. 2006; Ott et al. 1996; Pang et al. 2002; Preston and Mallory-Smith 2001; Schloss 1990; Singh and Shaner 1995).

4.1.2 ALS Gene Mutations That Confer Resistance to Herbicides

Single point mutations within multiple sites of the ALS gene can result in a variable pattern of cross-resistance between the classes of ALS-inhibiting herbicides (Shaner 1999). Initially these point mutations were characterized using mutants generated in the laboratory, e.g., ALS-inhibiting herbicide resistant tobacco (*Nicotiana tabacum* L.) cell cultures (Chaleff and Ray 1984; Creason et al. 1988; Hartnett et al. 1990; Lee et al. 1988), *Arabidopsis thaliana* (L.) Heynh. seeds (Haughn and Somerville 1986, 1990; Haughn et al. 1988; Mourad et al. 1992,

1993; Sathasivan et al. 1991), corn (*Zea mays* L.) cultures (Bernasconi et al. 1995), and *Brassica napus* L. cell cultures (Hattori et al. 1995). The same mutations were later detected in field-resistant plants, e.g., lettuce (*Lactuca serriola* L.) (Guttieri et al. 1992), kochia (*Kochia scoparia* (L.) Schrad.) (Guttieri et al. 1995; Saari et al. 1990), cocklebur (*Xanthium strumarium* L.) (Bernasconi et al. 1995), *Raphanus raphanistrum* L. (Boutsalis 2001; Hashem et al. 2001; Hashem and Bowran 2002; Tan and Medd 2002; Walsh et al. 2001), *Lindernia* (Itoh and Wang 1997; Itoh et al. 1999; Uchino and Watanabe 2002; Uchino et al. 1999, 2000), and *Amaranthus blitoides* S. Watson (Sibony and Rubin 1996, 2003). Biotypes in at least 20 monocotyledonous and 44 dicotyledonous plant species were recorded as having evolved resistance to several of the ALS-inhibiting herbicides (Heap 2003). This may be due to repeated applications of ALS inhibitor herbicides for more than three years (Rubin 1996). No reports for resistance of *Coreopsis tinctoria* Nutt. to imidazolinone or sulfonylurea herbicides have been previously published. The majority of mutations known to confer resistance to imidazolinone herbicides have been detected in domains A and B of the large subunit of the ALS gene (Tan et al. 2005; Wright et al. 1998). These mutations occur at positions (codons) Ala122, Pro197, Ala205, Trp574, and Ser653 (Amino acids numbered according to *A. thaliana* described in Sathasivan et al. 1990). Mutations at Ala122, Ser653, and Ala205 generally confer resistance to imidazolinones, but not sulfonlyureas (Tan et al. 2005). The most common mutations in biotypes selected by sulfonylureas occur in the highly conserved domain A with 13 amino acids, where any alteration of Pro197 confers resistance primarily to sulfonylureas and triazolopyrimiines (Guttieri et al. 1992). A Trp574 to Leu mutation in domain B has been associated with broad cross-resistance to representatives of all five families of ALS-inhibiting chemicals (Bernasconi et al. 1995; Tranel et al. 2006; Woodworth et al., 1996b). Some mutations occur in domain C where

an Ala122 to Thr mutation appears to confer resistance only to imidazolinones (Bernasconi et al. 1995), while an Ala205 to Val substitution in domain D confers broad cross-resistance (Woodworth et al. 1996a), as in the case of the Trp574 codon in Domain B.

In nearly all instances of enzyme-based resistance to ALS herbicides, resistance has been inherited as a single gene with varying degrees of dominance (Tranel and Wright 2002). Currie et al. (1995) demonstrated that ALS extracts from Pioneer IR corn hybrids were 6-fold more resistant to imazethapyr when compared to more than 62-fold resistance in homozygous plants. In the heterozygous XI-12 corn, imazethapyr resistance was 5-fold, compared to 250-fold in the homozygous plants, indicating that resistance in corn XI-12 is a semidominant trait (Wright and Penner 1998). Similar results were also obtained in *Sisymbrium orientale* L. (Boutsalis et al. 1999). Inheritance of resistance to imidazolinone or sulfonylurea herbicide has not been reported for *C. tinctoria*.

4.1.3 ALS-resistant Mutants in Crops and Weeds

More than 70 plant species throughout the world also have been reported showing resistance after repeated applications of ALS-inhibiting herbicides for more than three years (Heap 2003; Saari et al. 1994). Resistance in many cases has been attributed to single point mutations which can occur at multiple sites within the ALS gene (Shaner 1999) that provide opportunities to study the molecular basis of resistance and to transfer cloned resistance genes to different economic crops for weed management. Most ALS mutant lines, from laboratory or field sources, generally possess a nucleotide base-pair substitution at only one or two sites (Boutsalis et al. 1999), and do not always show a broad spectrum of resistance to ALS-inhibiting herbicides. Consequently, it would be beneficial to discover new gene sources of ALS that

exhibit a broad spectrum of resistance for basic biochemical and molecular studies and to transfer high levels of resistance to commercial crops.

Commercial varieties of rice (*Oryza sativa* L.), corn (*Zea mays* L.), oilseed rape (*Brassica napus* L.), sunflower (*Helianthus annuus* L.), and wheat (*Triticum aestivum* L.) that are resistant to imidazolinone herbicides have been released (Tan et al. 2005). Clearfield rice varieties CL121 and CL141, first released in 2001, were developed by chemical mutagenesis of seeds that created a single mutation at Gly654. Seeds from the variety Cypress were treated with the chemical EMS to produce a single mutation at Ser 653. The variety CL161 was developed from this mutation. Four single mutations in the corn ALS gene (codons Ala122, Ala155, Trp574, Ser653), produced from either cell culture or chemical mutagenesis, resulted in the release of "Clearfield corn" and other varieties. Oilseed rape mutants PM1 and PM2 were derived from microspore mutagenesis resulting in mutations at Ser653 and Trp574, respectively. "Clearfield" varieties of sunflower were produced from mutations at Ala205. Both winter and spring varieties of wheat have been released since 2001 that are resistant to imidazolinones due to a single mutation at Ser653. Mutations at Ala122, Pro197, and at other unknown sites were reported to confer resistance to imidazolinone herbicides in sugarbeet (*Beta vulgaris* L.), cotton (*Gossypium hirsutum* L.), soybean (*Glycine max* (L.) Merr.), lettuce (*Lactuca sativa* L.), tomato (*Lycopersicon esculentum* Mill.), and tobacco (*Nicotiana tabacum* L.).

A total of 23 herbicide-resistant weed species with mutations identified within the ALS gene have been posted online with periodical updates (http://www.weedscience.org/ mutations/ MutDisplay.aspx; Tranel, P.J., Wright, T.R, and Heap, I.M. ALS mutations from herbicide-resistant weeds. Online. Internet. Monday, January 23, 2006. Available http://www. weedscience.com). Mutations at Pro197 were the most common across 15 different species, 11

mutations at Trp574 in as many species were reported, and Ala122 and Ser653 were the least common in only 7 species. It is interesting that the mutation at Trp574, as was the case with commercial crops described above, resulted in high levels of resistance (> 10-fold) against all tested imidazolinone and sulfonylurea herbicides.

4.1.4 ALS Mutants of *Coreopsis tinctoria* Nutt.

No report has been published previously that describes ALS/AHAS-resistant mutants from *Coreopsis tinctoria* Nutt. Unlike most other ALS mutants produced from chemical mutagenesis, preliminary studies of the *Coreopsis tinctoria* Nutt. field-derived mutant plants suggest they are highly resistant to both sulfonylurea (sulfometuron methyl "Oust®"; chlorsulfuron "Glean®") and imidazolinone (imazapyr "Arsenal®") herbicides (our unpublished results).

4.1.5 Research Objectives

Due to patent applications by the Louisiana State University Agricultural Center and restrictions of patent applications, the specific objectives, except for the phylogenetic analysis, cannot be included in the dissertation at this time. The phylogenetic analysis objective is to: Perform phylogenetic analysis to determine relationship of *Coreopsis tinctoria* Nutt. ALS gene with 14 dicotyledonous and monocotyledonous plant species.

4.2 Materials and Methods

Due to patent applications, the Materials and Methods section, except for the phylogenetic analysis, cannot be included in the dissertation at this time.

GenBank was searched using the entire sequence of the mutant *Coreopsis tinctoria* L. ALS gene (sequence (1)). Fourteen other plant species were chosen from GenBank. The species were *Helianthus annuus* L. (GenBank accession AY541454 (1) and AY541457 (2)), *Xanthium*

strumarium L. (U16280), *Amaranthus powellii* S. Watson (AF363370), *Gossypium hirsutum* L. (Z46960), *Bassia scoparia* (L.) A. J. Scott (AF094326), *Arabidopsis thaliana* (L.) Heynh. (AY124092), *Nicotiana tabacum* L. (X07645), *Brassica napus* L. (Z11526), *Papaver rhoeas* L. (AJ577316), *Camelina microcarpa* Andrz. ex DC. (AY428947), *Zea mays* L. (X63553), *Oryza sativa* L. (AY885675), *Monochoria vaginalis* (Burm. f.) C. Presl ex Kunth (AB243613), and *Lolium multiflorum* Lam. (AF310684). DNA sequences were aligned using the program ClustalX (ftp://ftp-igbmc.u-strasbg.fr/pub/ ClustalX/), and a Neighbor Joining Tree was created using the program Mega3 (http://www.megasoftware.net).

4.3 Results

Due to patent applications, the Results section except for the phylogenetic analysis cannot be included at this time.

A search of the DNA sequence databases in the GenBank showed that the *Coreopsis tinctoria* Nutt. DNA sequence (1) contained significant sequence similarity to the higher plant ALS genes. Using the default parameters, the smallest sum probabilities ranged from 2.0×10^{-9} for *Lolium multiflorum* Lam. to 0 for *Helianthus annuus* L. Phylogenetic analysis based on the entire ALS gene sequences showed that the two different sequences of ALS gene from *Coreopsis tinctoria* Nutt. and the two different ALS genes of *Helianthus annuus* L. were most closely related, respectively (Fig. 4.1). The 15 different plant species were divided into four different groups (Fig. 4.1). Plants were monocots in group 1 and eudicots in group 2 to 4. Group 1 included *Lolium multiflorum* Lam., *Oryza sativa* L., and *Zea mays* L., members of *Poaceae* family and *Monochoria vaginalis* (Burm. f.) C. Presl ex Kunth, a member of *Commelinaceae* family. Group 2 contained *Papaver rhoeas* L., a member of *Ranunculaceae* family, and *Amaranthus powellii* S. Watson and *Bassia scoparia* (L.) A. J. Scott, members of

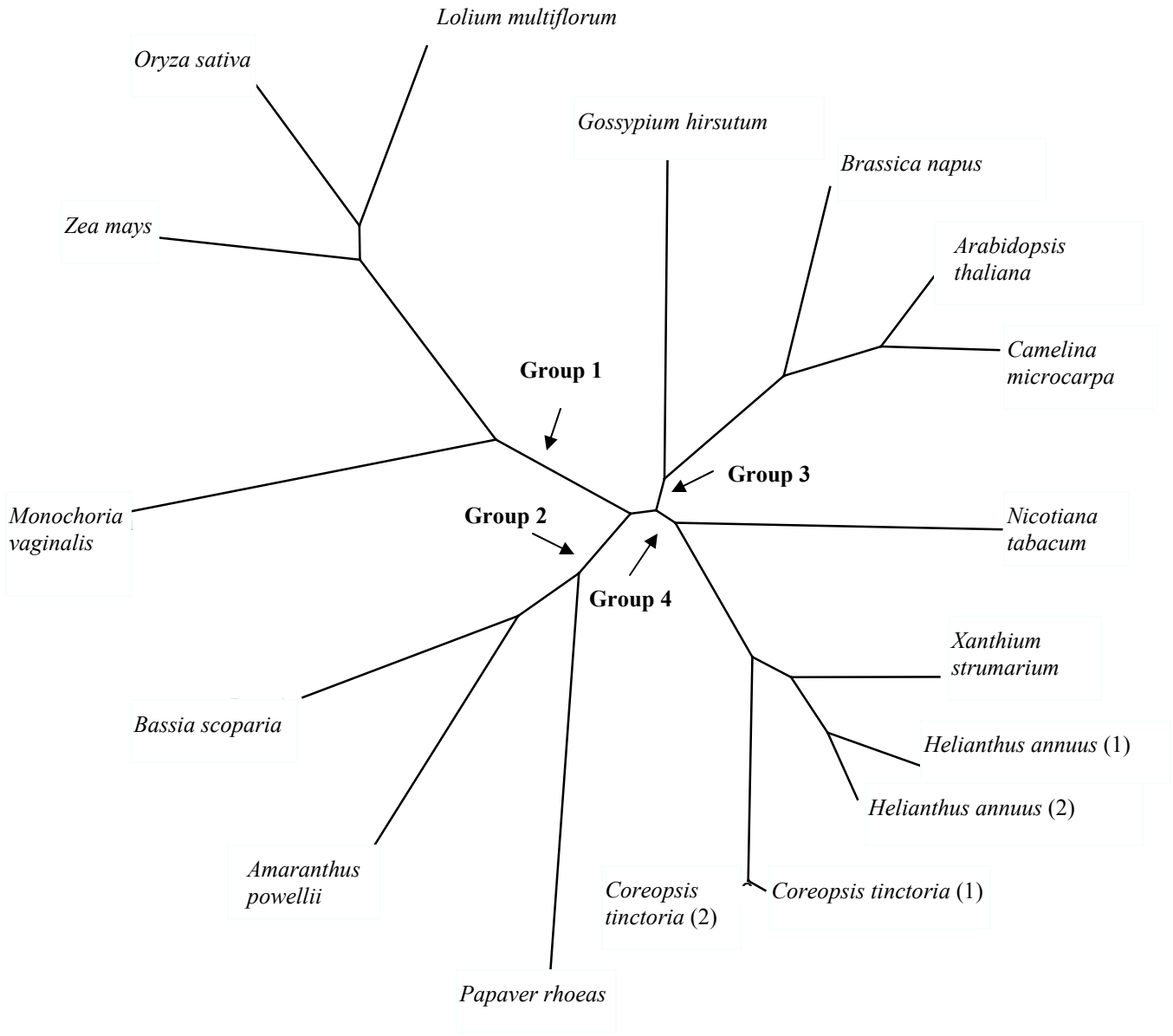

Fig. 4.1. Phylogenetic analysis of fifteen different plant species with the ALS gene.

Chenopodiaceae family. Group 3 had *Camelina microcarpa* Andrz. ex DC., *Arabidopsis thaliana* (L.) Heynh. and *Brassica napus* L., members of *Brassicaceae* family, and *Gossypium hirsutum* L., a member of *Malvaceae* family. *Nicotiana tabacum* L., a member of *Solanaceae* family, and *Coreopsis tinctoria* Nutt., *Xanthium strumarium* L. and *Helianthus annuus* L., members of *Asteraceae* family, belonged to Group 4. With the group number increasing, the relation magnitude of plant species among different groups was decreasing. Therefore, ALS of *Coreopsis tinctoria* Nutt. was most closely related to *Xanthium strumarium* L. and *Helianthus annuus* L. members of the *Asteraceae* family, and most distantly related to *Lolium multiflorum* Lam., *Oryza sativa* L., and *Zea mays* L., members of the *Poaceae* family. Plants of the *Asteraceae* family are known as dicots and those of the *Poaceae* family as monocots. Phylogenetic results obtained from the current study were consistent with traditional systematic analysis (angiosperm phylogeny website: http://www.mobot.org/MOBOT/Research/APweb/welcome.html).

4.4 Discussion

Due to patent applications, the Discussion section is restricted to the phylogenetic results in the dissertation at this time.

ALS genes were conserved across the different species of plants. The mutations occurred in the conserved region would cause plants to be tolerant to herbicides. Phylogenetic results based on ALS genes were consistent with traditional systematic analysis based on phenotypic data. Moreover, the fact that the different ALS sequences of same species were grouped together proved the reliability of the phylogenetic analysis. Therefore, the ALS genes will be a useful tool for future plant systematic analysis. If sampling more individual plants in each species and

sequencing the ALS genes for each individual plant, it will be possible to estimate the evolutionary profile of different species of plants.

4.5 References

Bekkaoui F., Schorr P., and Crosby W.L. 1993. Acetolactate synthase from *Brassica napus*: immunological characterization and quaternary structure of the native enzyme. Physiologia plantarum 88:475-484

Bernasconi P., Woodworth A.R., Rosen B.A., Subramanian M.V., and Siehl D.L. 1995. A naturally occurring point mutation confers broad range tolerance to herbicides that target acetolactate synthase. J. Biol. Chem. 270:17381-17385

Boutsalis P. 2001. Syngenta Quick-Test: a rapid whole-plant test for herbicide resistance. Weed Tech. 15:257-263

Boutsalis P., Karotam J., and Powles S.B. 1999. Molecular basis of resistance to acetolactate synthase-inhibiting herbicides in *Sisymbrium orientale* and Brassica *tournefortii*. Pesticide Sci. 55:507-516

Bowen T.L., Union J., Tumbula D.L., and Whitman W.B. 1997. Cloning and phylogenetic analysis of the genes encoding acetohydroxyacid synthase from the archaeon *Methanococcus aeolicus*. Gene 188:77-87

Chaleff R.S. and Ray T.B. 1984. Herbicide-resistant mutants from tobacco cell cultures. Science 223:1148-1151

Chaleff R.S. and Bascomb N.F. 1987. Genetic and biochemical evidence for multiple forms of actolactate synthase in *Nicotiana tabacum*. Mol. Gen. Genet. 210:33-38

Christopher J.T., Powles S.B., and Holtum J.A.M. 1992. Resistance to acetolactate synthase-inhibiting herbicides in annual ryegrass (*Lolium rigidum*) involves at least two mechanisms. Plant Physiol. 100:1909-1913

Christopher J.T., Powles S.B., Liljegren D.R., and Holtum J.A.M. 1991, Cross-resistance to herbicides in annual ryegrass (*Lolium rigidum*). Plant Physiol. 95 1036–1043

Cotterman J.C. and Saari L.L. 1992. Rapid metabolic inactivation is the basis for cross-resistance to chlorsulfuron in diclofop-methyl–resistant rigid ryegrass (*Lolium rigidum*) biotype SR4/84. Pestic. Biochem. Physiol. 43:182–192.

Creason G.L. and Chaleff R.S. 1988. A second mutation enhances resistance of a tobacco mutant to sulfonylurea herbicides. Theor. Appl. Genet. 76:177-182

Currie R.S., Kwon C.S., and Penner D. 1995. Magnitude of imazethapyr resistance of corn (*Zea mays*) hybrids with altered acetolactate synthase. Weed Sci. 43:578-582

Devine M.D., Marles M.A.S., and Hall L.M. 1991. Inhibition of acetolactate synthase in susceptible and resistant biotypes of *Stellaria media*. Pestic. Sci. 31:273-280

Duggleby R.G. and Pang S.S. 2000. Acetohydroxyacid synthase. J. Biochem and Mol. Biol. 33:1-36

Duggleby R.G., Pang S.S., Yu H., and Guddat L.W. 2003. Systematic characterization of mutations in yeast acetohydroxyacid synthase: interpretation of herbicide-resistance data. Eur. J. Biochem. 270:2894-2904

Falco S.C. and Dumas K.S. 1985. Genetic analysis of mutants of *Saccharomyces cerevisiae* resistant to the herbicide sulfometuron methyl. Genetics 109:21-35

Foes M.J., Liu L., Tranel P.J., Wax L.M., and Stoller E.W. 1998. A biotype of common waterhemp (*Amaranthus rudis*) resistant to triazine and ALS herbicides. Weed Science 46:514-520

Foes M.J., Liu L., Vigue G.E., Stoller W., Wax L.M., and Tranel P.J. 1999. A kochia (*Kochia scoparia*) biotype resistant to triazine and ALS-inhibiting herbicides. Weed Science 47:21-27

Grula J.W., Hudspeth R.L., Hobbs S.L., and Anderson D.M. 1995. Organization, inheritance and expressionof acetohydroxyacid synthase genes in the cotton allotetraploid *Gossypium hirsutum*. Plant Molec. Biol. 28:837-846

Guttieri M.J., Eberlein C.V., and Thill D.C. 1995. Diverse mutations in the acetolactate synthase gene confer chlorsulfuron resistance in kochia (*Kochia scoparia*) biotypes. Weed Sci. 43:175-182

Guttieri M.J., Eberlein C.V., Mallory-Smith C.A., and Thill D.C. 1996. Molecular genetics of target site resistance to acetolactate synthase inhibiting herbicides. In Molecular Genetics and ecology of Pesticide Resistance, T.M. Brown, ed (Washington, DC: American Chemical Society). pp10-16

Guttieri M.J., Eberlein C.V., Mallory-Smith C.A., Thill D.C., and Hoffman D.L. 1992. DNA sequence variation in domain A of the acetolactate synthase genes of herbicide resistant and susceptible weed biotypes. Weed Sci. 40:670-676

Hanson B.D., Park K.W., Mallory-Smith C.A., and Thill D.C. 2004. Resistance of *Camelina microcarpa* to acetolacate synthase inhibiting herbicides. Weed Research 44:187-194

Hartnett M.E., Chui C.F., Mauvais C.J., McDevitt R.E., Knowlton S., Smith J.K., Falco S.C., and Mazur B.J. 1990. Herbicide-resistant plants carrying mutated acetolactate synthase

genes. In managing resistance to agrochemicals, from fundamental research to practical strategies. M.B. Green, H.M. LeBaron, and W.K. Moberg, eds. ACS Symp. Ser. Am. Chem. Soc. 421, pp459-473

Hashem A. and Bowran D. 2002. Seed soaking method to test resistance in *Raphanus raphanistrum* to ALS-inhibiting herbicides. Plant. Prot. Qtly. 17:35-38

Hashem A., Bowran D., Piper T., and Dhammu H. 2001. Resistance of wild radish (*Raphanus raphanistrum*) to acetolactate synthase inhibiting herbicides in the Western Australia wheat belt. Weed Tech. 15:68-74

Hattori J., Brown D., Mourad G., Labbe H., Ouellet T., Sunohara G., Rutledge R., King J., and Miki B. 1995. An acetohydroxy acid synthase mutant reveals a single site involved in multiple herbicide resistance. Mol. Gen. Genet. 246:419-425

Haughn G.W. and Somerville C. 1986. Sulfonylurea-resistant mutants of *Arabidopsis thaliana*. Mol. Gen. Genet. 204:430-434

Haughn G.W. and Somerville C.R. 1990. A mutation causing imidazolinone resistance maps to *csr1* locus in *Arabidopsis thaliana* var Columbia. Plant Physiol. 92:1081-1085

Haughn G.W., Smith J., Mazur B., and Somerville C. 1988. Transformation with a mutant Arabidopsis acetolactate synthase gene renders tobacco resistant to sulfonylurea herbicides. Mol. Gen. Genet. 211:266-271

Heap I. 2003. International survey of herbicide resistant weeds. http://www.weedscience.com/

Hershey H.P., Schwartz L.J., Gale J.P., and Abell L.M. 1999. Cloning and functional expressionof the small subunit of acetolactate synthase from Nicotiana plumbaginifolia, Plant Mol. Biol. 40:795-806

Holtum J.A.M., Matthews J.M., Häusler R.E., Liljegren D.R., and Powles S.B. 1991. Cross resistance to herbicides in annual ryegrass (*Lolium rigidum*). III. On the mechanism of resistance to diclofop-methyl. Plant Physiol. 97:1026–1034

Itoh K. and Wang G.-X. 1997. An outbreak of sulfonylurea herbicide resistance in Scrophulariaceae paddy weeds in Japan. In Proceedings of the 16th Asian-Pacific Weed Science Society Conference (Kuala Lumpur, Malaysia, 8-12 September 1997) (Kuala Lumpur: Malaysian Plant Protection Society), pp219-221

Itoh K., Wang G.-X., and Ohba S. 1999. Sulfonylurea resistance in *Lindernian micrantha*, an annual paddy weed in Japan. Weed Res. 39:413-423

Jander G., Baerson S.R., Hudak J.A., Gonzalez K.A., Gruys K.J., and Last R.L. 2003. Ethylmethanesulfonate Saturation Mutagenesis in Arabidopsis to Determine Frequency of Herbicide Resistance. Plant Physiol. 131:139-146

Kemp M.S., Moss S.R., and Thomas T.H. 1990. Herbicide resistance in Alopecurus myosuroides. In: Green M.B., LeBaron H.M., and Moberg W.K. (eds) Managing resistance to agrochemicals. From fundamental research to practical strategies. Washington D.C.: American Chemical Society. pp376-393

Kolkman J.M., Slabaugh M.B., Bruniard J.M., Berry S., Bushman B.S., Olungu C., Maes N., Abratti G., Zambelli A., Miller J.F., Leon A., and Knapp S.J. 2004. Acetohydroxyacid synthase mutations conferring resistance to imidazolinone or sulfonylurea herbicides in sunflower. Theor. Appl. Genet. 109:1147-1159

Kovar J.L., Zhang J., Funke R.P., and Weeks D.P. 2002. Molecular analysis of the acetolactate synthase gene of *Chlamydomonas reinhardtii* and development of a genetically engineered gene as a dominant selectable marker for genetic transformation. The Plant Journal 29:109-117

LaRossa R.A. and Schloss J.V. 1984. The sulfonylurea herbicide sulfometuron methyl is an extremely potent and selective inhibitor of acetolactate synthase in *Salmonella typhimurium*. J. Biol. Chem. 259:8753-8757

Lawson M.J. and Zhang L. 2006. Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes. Genome Biology 7:R14

Lee K.Y., Townsend J., Tepperman J., Black M., Chui C.F., Mazur B., Dunsmuir P., and Bedbrook J. 1988. The molecular basis of sulfonylurea herbicide resistance in tobacco. EMBOJ 7:1241-1248

Lee Y. and Duggleby R.G. 2002. Regulatory interactions in *Arabidopsis thaliana* acetohydroxyacid synthase. FEBS letters 512:180-184

Levitt G. 1978. Herbicide sulfonamides. US Patent 4127405

Mallory-Smith C.A. and Retzinger E.J. 2003. Revised classification of herbicides by site of action of weed resistance management strategies. Weed Tech. 17:605-619

Manley B.S., Singh B.K., Shaner D.L., and Wilson H.P. 1999. Imidazolinone resistance in smooth pigweed (*Amaranthus hybridus*) is due to an altered acetolactate synthase. Weed Tech. 13:697-705

Mazur B.J. and Falco S.C. 1989. The development of herbicide resistant crops. Ann. Rev. Plant Physiol. Plant Mol. Biol. 40:441-470

Mazur B.J., Chui C.-F., and Smith J.K. 1987. Isolation and characterization of plant genes coding for acetolactate synthase, the target enzyme for two classes of herbicides. Plant Physiol. 85:1110-1117

McCourt J.A., Pang S.S., King-Scott J., Guddat L.W., and Duggleby R.G. 2006. Herbicide-binding sites revealed in the structure of plant acetohydroxyacid synthase. PNAS 103:569-573

McNaughton K.E., Lee E.A., and Tardif F.J. 2001. Mutations in the ALS gene conferring resistance to group II herbicides in redroot pigweed (*Amaranthus retroflexus*) and green pigweed (*A. powellii*). Weed Sci. Soc. Am. Abstr. 41:97

Moss S.R. and Cussans G.W. 1991. The development of herbicide resistant populations of *Alopecurus myosuroides* (black-grass) in England. In Caseley J.C., Cussans G.W., and Atkin R.K. (eds). Herbicide resistance in weeds and crops. Oxford, U.K.: Butterworth-Heneman pp445-456

Mourad G. and King J. 1992. Effect of four classes of herbicides on growth and acetolactate-synthase activity in several variants of *Arabidopsis thaliana*. Planta 188:491-497

Mourad G., Pandey B., and King J. 1993. Isolation and genetic analysis of a triazolopyrimidine-resistant mutant of *Arabidopsis*. J. Hered. 84:91-96

Newhouse k., Singh B., Shaner D., and Stidham M. 1991. Mutations in corn (Zea mays L.) conferring resistance to imidazolinone herbicides. Theor. Appl. Genet. 83:65-70

Ott K., Kwagh J., Stockton G.W., Sidorov V., and Kakefuda Gl, 1996. Rational molecular design and gnentic engineering of herbicide resistant crops by structure modeling and site-directed mutagenesis of acetohydroxyacid synthase. J. Mol. Biol. 263:359-368

Pang S.S., Duggleby R.G., and Guddat L.W. 2002. Crystal structure of yeast acetohydroxyacid synthase: a target for herbicidal inhibitors. J. Mol. Biol. 317:249-262

Park K.W. and Mallory-Smith C.A. 2004. Physiological and molecular basis for ALS inhibitor resistance in *Bromus tectorum* biotypes. Weed Research 44:71-77

Patzoldt W.L. and Tranel P.J. 2001. ALS mutations conferring herbicide resistance in waterhemp. Proceedings of the North Central Weed Science Society 56:67

Patzoldt W.L. and Tranel P.J. 2002. Molecular analysis of cloransulam resistance in a population of giant ragweed. Weed Science 50:299-305

Patzoldt W.L., Tranel P.J., Alexander A.L., and Schmitzer P.R. 2001. A common ragweed population resistant to cloransulam-methyl. Weed Science 49:485-490

Preston C. and Mallory-Smith C.A. 2001. Biochemical mechanisms, inheritance, and molecular genetics of herbicide resistance in weeds. In Herbicide resistance and world grains, Powles S.B. and Shaner D.L., eds (Boca Raton, FL: CRC Press), pp23-60

Ray T.B. 1984. Site of action of chlorsulfuron: inhibition of valine and isoleucine biosynthesis in plants. Plant Physiol. 75:827-831

Rubin B. 1996. Herbicide-resistant weeds — the inevitable phenomenon: mechanisms, distribution and significance. Z Pflanzenkr Pflanzenschutz XV, 17-32

Rutledge R., Ouellet T., Hattori J., and Miki B.L. 1991. Molecular characterization and genetic origin of the *Brassica napus* acetohydroxyacid synthase multigene family. Mol. Gen. Genet. 229:31-40

Saari L.L., Cotterman J.C., and Primiani M.M. 1990. Mechanism of sulfonylurea herbicide resistance in the broadleaf weed, *Kochia scoparia*. Plant Physiol. 93, 55-61

Saari L.L., Cotterman J.C., and Thill D.C. 1994. Resistance to acetolactate synthase inhibiting herbicides. In Herbicide Resistance in Plants: Biology and Biochemistry, S.B. Powles and J.A.M. Holtum, eds (Boca Raton, FL: CRC Press), pp83-139

Saari L.L., Cotterman J.C., Smith W.F., and Primiani M.M. 1992. Sulfonylurea herbicide resistance in common chickweed, perennial ryegrass, and Russian thistle, Pestic. Biochem. Physiol. 42:110-118

Santel H.J., Bowden B.A., Sorenson V.M., Mueller K.H., and Reynolds J. 1999. Flucarbazone-sodium—a new herbicide for grass control in wheat. Proc. West. Soc. Weed Sci. 52:124–125

Sathasivan K., Haughn G.W., and Murai N. 1990. Nucleotide sequence of a mutant acetohydroxyacid synthase gene from an imidazolinone-resistant *Arabidopsis thaliana* var. Columbia. Nucleic Acids Res. 18:2188

Sathasivan K., Haughn G.W., and Murai N. 1991. Molecular basis of imidazolinone herbicide resistance in *Arabidopsis thaliana* var Columbia. Plant Physiol. 97:1044-1050

Scarabel L., Carraro N., Sattin M., and Varotto S. 2004. Molecular basis and genetic characterisation of evolved resistance to ALS-inhibitors in *Papaver rhoeas*. Plant Science 166: 703-709

Scarabel L., Varotto S., and Sattin M. 2005. An italian population of *Amaranthus retroflexus* resistant to ALS-inhibiting herbicides: resistance pattern and molecular basis. 13th EWRS Symposium Abstr. 202

Schloss J.V. 1990. Acetolactate synthase, mechanism of action and its herbicide binding site. Pestic. Sci. 29:283-292

Schmenk R.E., Barrett M., and Witt W.E. 1997. An investigation of smooth pigweed (*Amaranthus hybridus* L.) resistance to acetolactate synthase inhibiting herbicides. Weed Science Society of America Abstracts 37:117

Shaner D.L. 1991. Physiological effects of the imidazolinone herbicides. In: Shaner D.L. and O'Connor S.L. (eds) The imidazolinone herbicides. Lews, Ann Arbor, Mich., pp129-138

Shaner D.L. 1999. Resistance to acetolactate synthase (ALS) inhibitors in the United States: history, occurrence, detection and management. J. Weed Sci. Technol. 44: 405-411

Shaner D.L., Anderson P.C., and Stidham M.A. 1984. Imidazolinones: potent inhibitors of acetohydroxyacid synthase. Plant. Phsiol. 76:545-546

Sibony M. and Rubin B. 1996. Multiple resistance to ALS inhibitors and triazine herbicides in *Amaranthus* species. In Proceedings of the international symposium on weed and crop resistance to herbicides (Cordoba, Spain), pp112-114

Sibony M. and Rubin B. 2003. Molecular basis for multiple resistance to acetolactate synthase-inhibiting herbicides and atrazine in *Amaranthus blitoides* (prostrate pigweed). Planta 216:1022-1027

Sibony M., Michel A., Haas H.U., Rubin B., and Hurle K. 2001. Sulfometuron-resistant *Amaranthus retroflexus*: cross-resistance and molecular basis for resistance to acetolactate synthase-inhibiting herbicides. Weed Research 41:509-522

Singh B.K. and Shaner D.L. 1995. Biosynthesis of branched chain amino acids: from test tube to field. Plant Cell 7:935-944

Singh B.K., Stidham M.A., and Shaner D.L. 1988. Assay of acetohydroxyacid synthase. Analytical Biochemistry 171:173-179

Stidham M.A. and Shaner D.L. 1990. Imidazolinone inhibition of acetohydroxyacid synthase in vitro and in vivo. Pestic. Sci. 29:335-340

Strother J.L. 1983. More chromosome studies in Compositae. Amer. J. Bot. 70:1217-1224

Subramanian M.V. and Gerwick B.C. 1989. Inhibition of acetolactate synthase by triazolopyrimidines. *In* J.R. Whitaker et. al. (ed.) Biocatalysis in agricultural biotechnology. ACS Symposium Series. American Chemical Society, Washington, DC. pp277–288

Subramanian M.V., Hung H.Y., Dias J.M., Miner V.W., Butler J.H., and Jachetta J.J. 1990. Properties of mutant acetolactate synthases resistant to triazolopyrimidine sulfonanilide. Plant Physiol. 94:239-244

Tal A. and Rubin B. 2004. Occurrence of resistant Chrysanthemum coronarium to ALS inhibiting herbicides in Israel. Resistant Pest Management Newsletter 13:31-33

Tan M.K. and Medd R.W. 2002. Characterization of the acetolactate synthase (ALS) gene of *Raphanus raphanistrum* L. and the molecular assay of mutations associated with herbicide resistance. Plant Science 163:195-205

Tan S., Evans R.R., Dahmer M.L., Singh B.K., and Shaner D.L. 2005. Imidazolinone-tolerant crops: history, current status and future. Pest Manag. Sci. 61:246-257

Thill D.C., Mallory-Smith C.A., Saari L.L., Cotterman J.C., Primiani M.M., and Saladini J.L. 1993. Sulfonylurea herbicide resistant weeds: discovery, distribution, biology, mechanism, and management. In Herbicide Resistance in Weeds and Crops, J.C. Caseley, G.W. Cussans, and R.K. Atkin, eds (London: Butterworth and Heinemann), pp115-128

Tranel P.J. and Wright T.R. 2002. Resistance of weeds to AHAS-inhibiting herbicides: what have we learned? Weed Sci 50:700-712

Tranel P.J., Wright T.R, and Heap, I.M. 2006. ALS mutations from herbicide-resistant weeds. Online. Internet. Sunday, March 12, 2006. Available http://www.weedscience.com.

Tu M., Hurd C., and Randall J.M. 2001. Weed Control Methods Handbook: Tools and Techniques for Use in Natural Areas. In: http://tncweeds.ucdavis.edu/handbook.html

Uchino A. and Watanabe H. 2002. Mutations in the acetolactate synthase genes of sulfonylurea-resistant biotypes of *Lindernia spp*. Weed Biology and Management 2:104-109

Uchino A., Itoh K., Wang G.-X., and Tachibana M. 2000. Sulfonylurea resistant biotypes of *Lindernia* species in the Tohoku region and their response to several herbicides. J. Weed Sci. Technol. 45:13-20

Uchino A., Watanabe H., Wang G.-X., and Itoh K. 1999. Light requirement in rapid diagnosis of sulfonylurea-resistant weeds of *Lindernia* spp. (Scrophulariaceae). Weed Technol. 13:680-684

Veldhuis L.J., Hall L.M., O'Donovan J.T., Dyer W., and Hall J.C. 2000. Metabolism-based resistance of a wild mustard (*Sinapis arvensis* L.) biotype to ethametsulfuron-methyl. J. Agric. Food Chem. 48:2986-2990

Walsh M.J., Duane R.D., and Powles S.B. 2001. High frequency of chlorsulfuron resistant wild radish (*Raphanus raphanistrum* L.) populations across the Western Australian wheatbelt. Weed Tech. 15:199-203

Woodworth A.R., Bernasconi P., Subramanian M.V., and Rosen B.A. 1996a. A second naturally occurring point mutation confers broad-based tolerance to acetohydroxyacid synthase inhibitors. Plant Physiol. 111S:415 (Abstract)

Woodworth A.R., Rosen B.A., and Bernasconi P. 1996b. Broad range resistance to herbicides targeting acetohydroxyacid synthase (ALS) in a field isolate of *Amaranthus* sp. is conferred by a Try to Leu mutation in the ALS gene (Accession Number U55852) (PGR96-051). Plant Physiol. 111:1353

Wright T.R. and Penner D. 1998. Corn (Zea mays) acetolactate synthase sensitivity to four classes of ALS-inhibiting herbicides. Weed Sci. 46:8-12

Wright T.R., Bascomb N.F., Sturner S.F., and Penner D. 1998. Biochemical mechanism and molecular basis for ALS-inhibiting herbicide resistance in sugarbeet (*Beta vulgaris*) somatic cell selections. Weed Sci. 46:13-23

Yu Q., Zhang X.Q., Hashem A., Walsh M.J., and Powles S.B. 2003. ALS gene proline (197) mutations confer ALS herbicide resistance in eight separated wild radish (*Raphanus raphanistrum*) populations. Weed Science 51:831-838

CHAPTER 5: SUMMARY AND CONCLUSIONS

The first research objective of this dissertation described in chapter two dealt with development and evaluation of several statistical procedures to identify DNA markers associated with economically important breeding traits among a diverse set of inbred rice lines. This general area of research is referred to as association genetics and is different in structure and analysis when compared to the standard genetic mapping approach of backcross or recombinant inbred populations. Most advances in association genetics were historically driven by human geneticists, but several statistical models proposed by Ed Buckler's lab (discussed below) at Cornell and results shown in chapter two have contributed alternative approaches to marker identification of complex traits in plants.

The justification for this research is related to potential advantages of association genetics over traditional mapping approaches that first require marker screens of parents of a cross for polymorphic loci and then development of a recombinant inbred population of ≥ 300 lines to create a linkage map that takes three to fours years to develop. Moreover, statistical power and precision are limited in a diploid species like rice due to reduced meiotic events and segregation of only two alleles at a single locus for any given population. Large segregating or intermating populations can be developed to overcome these limitations, but few researchers are willing or able to bear such financial and labor commitments on a routine basis.

Association genetics has the potential to overcome the above-mentioned constraints by capturing multiple historical recombination events embedded within a diverse set of inbred lines such as those evaluated in chapter two. A "novel" approach for plant association genetic studies of 12 agronomic traits described in this dissertation is based on a multivariate procedure referred to as discriminant analysis (DA). This method uses stepwise discriminant analysis to identify

markers in two or more subsets of the population referred to as "training samples" that best differentiate the subsets. This step is crucial to reduce the number of markers evaluated that in turn diminishes Type I errors due to repeated measures using a large number of markers. Results from this study, published in the journal *Theoretical and Applied Genetics* (Zhang et al. 2005), indicate that DA-selected markers for the 12 traits mapped to the same or nearby genetic regions when compared to previous studies by various researchers that used traditional mapping techniques. New markers identified by DA suggest that the procedure can also uncover relevant genetic regions not possible by standard genetic tests. The DA-based approach has been recently used by the Louisiana State University Agricultural Center sweet potato breeding program (Dr. Labonte) to successfully identify DNA markers associated with disease resistance and dry matter yield. Taken together, the seven-step procedure of association genetics described in chapter two should be considered as a viable approach to identify candidate markers for simple and complex agronomic traits that have a direct and immediate impact on rice germplasm and varietal development.

Current statistical methods for association genetics, including the DA-based approach described above are based on the "general linear model" (GLM) where DNA markers used as predictor variables are assumed to act independently with no interaction effects. A popular software program for plant GLM analysis is the "TASSEL" method developed by Ed Buckler's laboratory at Cornell University. Of course, the assumption of locus independence is not realistic for complex regulatory loci that interact and participate in cascading signaling pathways. To overcome these limitations, an epistatic model described in chapter three, was developed using the newly released SAS GLMSELECT procedure. In addition, two "new" GLM models referred

to as the "LASSO" method of Tibshirani (1996) and LAR method of Efron et. al. (2004) were also evaluated to reduce parameter dimension and increase precision.

Results from the study shown in chapter three indicate that the epistatic model based on stepwise forward regression did successfully identify several interacting loci that explained a relatively high proportion of the observed variation for all the twelve agronomically important traits. Moreover, the loci identified by the epistatic model mapped within previously known QTL regions that underscores the genetic basis of the selected markers. It was concluded that stepwise forward regression with consideration for population structure, epistatic interactions, and missing data (multiple imputation) was a robust method, compared to the general linear model, to identify markers associated with complex agronomic traits. A 0.05 significance level and the smallest Predicted Residual Sum of Squares statistic of Cross Validation (CVPRESS) were used as criteria for optimal selection of both additive and epistatic models. Additional selection criteria should be evaluated in future studies. The selected interacting loci are of particular genetic and breeding interest that merit further investigation.

The second research objective was focused on genetic analysis of herbicide resistance from mutants of *Coreopsis tinctoria*. In this study the mutants were compared by chemical, genetic, and molecular analyses with "normal" or wild-type *Coreopsis.* Due to restrictions of patent applications, specific results from this research cannot be presented at this time. However, it can be stated that results from this study would have a direct and immediate beneficial impact on weed control strategies for agricultural plant commodities in Louisiana.

APPENDIX: LETTER OF PERMISSION

From: "Essenpreis, Alice, Springer DE" <Alice.Essenpreis@springer.com>
To: nzhang1@lsu.edu
CC:
Subject: WG: Permission to use in dissertation
Date: Mon, 12 Jun 2006 10:19:50 +0200

Dear Nengyi Zhang,

With reference to your request (copy herewith) to re-use material on which Springer controls the copyright, our permission is granted free of charge, on the following conditions:

* it concerns original material which does not carry references to other sources,
* if material in question appears with credit to another source, authorization from and reference to that source is required as well, and permission is also obtained from the author (address is given on the imprint page or with the article);

* allows you non-exclusive reproduction rights throughout the world,
* permission includes use in an electronic form, on the condition that content is
  - password protected,
  - at Intranet or
  - in CD-ROM/E-book;
* full credit (book/journal title, volume, year of publication, page, chapter/article title, name(s) of author(s), figure number(s), original copyright notice) is given to the publication in which the material was originally published by adding: With kind permission of Springer Science and Business Media.

Permission free of charge does not prejudice any rights we might have to charge for reproduction of our copyrighted material in the future.

With best regards,

Alice Essenpreis
Springer
Rights and Permissions

Tiergartenstrasse 17 | 69121 Heidelberg GERMANY
FAX: +49 6221 487 8223
Alice.Essenpreis@springer.com
WWW.springer.com/rights

Von: Nengyi Zhang [mailto:nzhang1@lsu.edu]
Gesendet: 30 May 2006 22:19
An: Schreck, Sabine, Springer DE
Betreff: Permission to use in dissertation

Dear Sir/Madam:

I wish to request permission to use the publication cited below as a part of my dissertation. I am a Ph.D. student working under Dr. James Oard in Louisiana State University and the senior author of the paper.

Zhang N., Xu Y., Akash M., McCouch S., and Oard J.H. 2005. Identification of Candidate Markers Associated with Agronomic Traits in Rice Using Discriminant Analysis. Theor Appl Genet 110:721-729

DOI 10.1007/s00122-004-1898-z

Thank you very much!

Sincerely,

Nengyi Zhang

VITA

Nengyi Zhang was born at Kaihua County, Zhejiang Province of China, on June 15, 1969. He attended Kaihua High School from 1984 to 1987. After graduating from high school, he attended Hangzhou University, China, from 1987 to 1991 for his Bachelor of Science degree in biology. In 1991, he attended Zhejiang Agricultural University for his Master of Science degree in plant breeding and genetics. After completing his master's degree in 1994, he worked in the Zhejiang Agricultural University from 1994 to 1998 and Zhejiang University[1] from 1998 to 2000 as Rice Breeder, Lecturer and Associate Director of the Division of Crop Genetics and Breeding, Agronomy Department. He taught undergraduate students genetics and plant breeding and graduate students cell biology and advanced plant genetics and breeding. He developed three released *Japonica* varieties, and one released hybrid rice variety. From 2000 to 2002, he worked at Louisiana State University as a visiting scholar, conducting research on risk assessment of transgenic herbicide technology in rice.

Nengyi was granted the Fellowship of Biotechnology Education for Students and Teachers (BEST) provided by Louisiana State University Agricultural Center for his doctoral degree study in plant genetics and molecular biology in the Department of Agronomy and Environmental Management, Louisiana State University, in August 2002, which he completed in the Summer of 2006. His doctoral research included association genetics for agronomic traits in rice and cloning of ALS herbicide resistant genes from *Coreopsis tinctoria* Nutt. Nengyi is a recipient of the Gerald O. Mott Meritorious Award in Crop Science offered by the Crop Science Society of America (CSSA). He has more than 20 publications in different journals.

[1] Since September 15, 1998, the four universities, Zhejiang University, Hangzhou University, Zhejiang Agricultural University, and Zhejiang Medical University have been merged into one university, Zhejiang University.